Environment Canada Environnement Canada

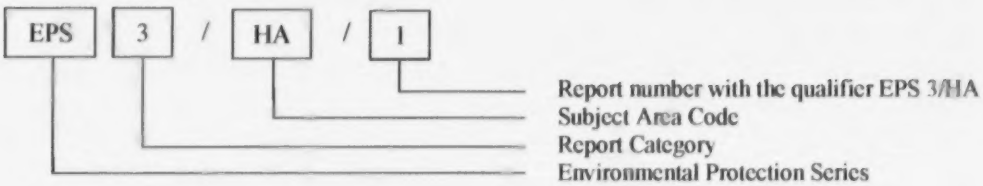
Guidance Document on the Sampling and Preparation of Contaminated Soil for Use in Biological Testing

EPS 1/RM/53 – February 2012
Science and Technology Branch
Environment Canada

Canada

Environmental Protection Series

Sample Number:



Categories

1 Regulations/Guidelines/Codes of Practice
2 Problem Assessments and Control Options
3 Research and Technology Development
4 Literature Reviews
5 Surveys
6 Social, Economic and Environmental Impact Assessments
7 Surveillance
8 Policy Proposals and Statements
9 Manuals

Subject Areas

AG	Agriculture
AN	Anaerobic Technology
AP	Airborne Pollutants
AT	Aquatic Toxicity
CC	Commercial Chemicals
CE	Consumers and the Environment
CI	Chemical Industries
FA	Federal Activities
FP	Food Processing
HA	Hazardous Wastes
IC	Inorganic Chemicals
MA	Marine Pollutants
MM	Mining and Ore Processing
NR	Northern and Rural Regions
PF	Paper and Fibres
PG	Power Generation
PN	Petroleum and Natural Gas
RA	Refrigeration and Air Conditioning
RM	Reference Methods
SF	Surface Finishing
SP	Oil and Chemical Spills
SRM	Standard Reference Methods
TS	Transportation Systems
TX	Textiles
UP	Urban Pollution
WP	Wood Protection/Preservation

New subject areas and codes are introduced as they become necessary. A list of EPS reports may be obtained from Communications Services, Environment Canada, Ottawa, Ontario, K1A 0H3, Canada.

Guidance Document on the Sampling and Preparation of Contaminated Soil for Use in Biological Testing

Method Development and Applications Unit
Science and Technology Branch
Environment Canada
Ottawa, Ontario

EPS 1/RM/53
February 2012

Cat. No.: En49-7/1-53E
ISBN: 978-1-100-19576-6

Aussi disponible en français

Reader's Comments

Comments regarding the content of this report should be addressed to:

Richard P. Scroggins, Chief
Biological Assessment and Standardization Section
Science and Technology Branch
Environment Canada
335 River Road
Ottawa, Ontario
K1A 0H3

Lisa Taylor, Manager
Method Development and Applications Unit
Science and Technology Branch
Environment Canada
335 River Road
Ottawa, Ontario
K1A 0H3

Cette publication est aussi disponible en français. Pour l'obtenir, s'addresser à :

Publications de la Protection de l'environnement
Environnement Canada
Ottawa (Ontario)
K1A 0H3

Review Notice

This report has been reviewed by staff from the Science and Technology Branch, Environment Canada, and approved for publication. Mention of trade names or commercial products does not constitute endorsement by Environment Canada for use. Other products of similar value are available.

Abstract

This document provides information for contaminated land site assessors, risk assessors, site managers, and risk managers on the design and implementation of a standardized, best practices approach for sampling and preparing soil for single-species toxicity and microbial tests. The guidance is also applicable for multi-species microcosm tests. Detailed procedures on the sampling, handling, transport, storage, and preparation of contaminated and reference soil for use in biological testing and complements are provided, but these do not replace the guidance that already exists for the sampling, collection, handling, and preparation of soils for chemical analyses.

Site-specific biological testing is an important component of contaminated land assessments as the data from site-specific biological testing integrate the inherent toxicity of the contaminant(s) and their mixtures in the soil with the bioavailability of the contaminant(s) under the specific conditions found at the site. Numerous biological tests are recommended in this document for the assessment of the toxicity of contaminated land including those for plants, macroinvertebrates, mesofauna, microorganisms, and microbial-mediated processes. Guidance is provided for when it is appropriate to use these site-specific biological tests, ranging from the screening level stage to use in higher-tier risk assessments. Case studies describing how biological test data are used in contaminated land management are also provided.

General or universal procedures are outlined for the preparation of collecting soil samples and include: developing study objectives; identifying the study area; collecting background data; conducting site surveys, soil surveys and ecological land classifications; selecting sampling strategies and locations; determining the size and number of samples to collect; establishing proper quality assurance and quality control procedures; considering environment, health and safety; and, developing sampling plans. General or universal procedures are also provided for: selecting sampling devices; collecting soil samples by horizon or by depth; handling soil samples on-site; selecting sample containers; and, transporting samples. There are general or universal procedures for personnel receiving, preparing and storing soil samples for biological testing, and soil preparation procedures described include drying, wetting, sieving, grinding, homogenizing, reconstituting, and characterizing samples. Examples of sampling devices, containers, and paper forms are provided in appendices. Specific examples of how to select sampling strategies and calculate the number of samples to collect are also provided as case studies in an appendix.

Additional procedures and considerations are included that are specific to the nature of the soil contaminant(s), biological testing requirements, and study objectives. These include procedures for collecting, handling, and preparing soils contaminated with volatile or unstable contaminants and manipulating soil samples (e.g., washing soils, autoclaving soils, etc.). An additional section is dedicated to guidance for sampling, handling, transporting, storing, and preparing soil from Canada's largest ecozones, boreal forest, taiga and tundra, as well as organic and wetland soils, for which some of the universal collection, handling, transport, storage and preparation procedures require modification. A brief description of ecological land classification, soil classification, and descriptions of the most dominant Soil Orders in Canada are provided in appendices.

Résumé

Le présent document fournit aux évaluateurs et gestionnaires de lieux contaminés, de même qu'aux évaluateurs et gestionnaires du risque, de l'information sur la conception et l'application de pratiques exemplaires normalisées d'échantillonnage et de préparation de sol aux fins d'essais toxicologiques monospécifiques et d'essais sur des organismes microbiens. Le guide peut aussi servir à des essais plurispécifiques en microcosme. Les procédures détaillées qui y sont décrites ont trait à l'échantillonnage, la manipulation, le transport, l'entreposage et la préparation de sol contaminé et de sol de référence à utiliser dans les essais biologiques et leurs compléments. Toutefois, ces procédures ne remplacent pas celles qui existent déjà pour l'échantillonnage, le prélèvement, la manipulation et la préparation de sol aux fins d'analyses chimiques.

Les essais biologiques propres à un site constituent un élément important de l'évaluation des terrains contaminés, car les données qui en sont tirées permettent d'établir des liens entre la toxicité intrinsèque du contaminant (il peut y en avoir plus d'un) et de ses mélanges dans le sol, d'une part, et la biodisponibilité du contaminant dans les conditions propres au site, d'autre part. Le présent guide recommande de nombreux essais biologiques applicables à l'évaluation de la toxicité d'un terrain contaminé, dont des essais qui font appel à des plantes, des macro-invertébrés, des espèces mésofauniques et des micro-organismes ou qui ont trait à des processus à médiation microbienne. Des indications portent sur la pertinence de chacun des essais biologiques propres à un site, qui vont de l'analyse préalable à des évaluations du risque à des niveaux plus élevés. Le guide renferme également des études de cas sur l'utilisation de données d'essais biologiques dans la gestion de terrains contaminés.

Les modes opératoires généraux ou universels applicables aux préparatifs entourant le prélèvement d'échantillons de sol incluent les suivants : établissement des objectifs de l'étude; délimitation de la zone d'étude; collecte de données documentaires; levés du site, levés pédologiques et classification écologique du sol; choix des stratégies et des lieux d'échantillonnage; détermination du nombre et de la taille des échantillons à prélever; établissement de procédures adéquates d'assurance et de contrôle de la qualité; facteurs à prendre en compte en matière d'environnement, de santé et de sécurité; conception de plans d'échantillonnage. D'autres modes opératoires généraux ou universels ont trait au choix des échantillonneurs, au prélèvement d'échantillons de sol selon l'horizon ou la profondeur, à la manipulation des échantillons in situ, au choix des récipients à échantillon et au transport des échantillons. Enfin, des modes opératoires généraux ou universels sont fournis à l'intention du personnel chargé de recevoir, de préparer et d'entreposer les échantillons de sol aux fins des essais biologiques. Les procédures de préparation du sol incluent l'assèchement, l'humidification, le tamisage, le broyage, l'homogénéisation, la reconstitution et la caractérisation des échantillons. Les annexes renferment des exemples d'échantillonneurs et de récipients, de même que des formulaires. On y trouve aussi, sous forme d'études de cas, des exemples précis sur la façon de choisir les stratégies d'échantillonnage et de calculer le nombre d'échantillons à prélever.

Le présent document précise la marche à suivre et les points à examiner en fonction de la nature du ou des contaminants du sol, des exigences des essais biologiques et des objectifs de l'étude, notamment en ce qui a trait au prélèvement, à la manipulation et à la préparation de sols renfermant des contaminants volatils ou instables, de même qu'au conditionnement des échantillons de sol (p. ex., lavage et autoclavage). Une section supplémentaire renferme des indications sur l'échantillonnage, la manipulation, le transport, l'entreposage et la préparation de sol provenant des plus grandes écozones du Canada, de la forêt boréale, de la taïga et de la toundra. Dans le cas des sols organiques et des milieux humides, certains des modes opératoires universels applicables à leur prélèvement, leur manipulation, leur transport, leur entreposage et leur préparation exigent des modifications. Les annexes renferment également une brève description de la classification écologique des terres, de la classification des sols et des plus importants ordres de sols du Canada.

Foreword

A series of guidance manuals and recommended test methods for measuring and assessing the biological effects of toxic substancesor materials in aquatic and terrestrial environments has been developed by Environment Canada.

Recommended guidance methods are those which have been evaluated by the Environmental Protection Service (EPS) and are favoured:

- *for use in Environment Canada environmental toxicology laboratories;*
- *for testing which is contracted out by Environment Canada or requested from outside agencies or industry;*
- *in the absence of more specific instructions, such as are contained in regulations; and,*
- *as a foundation for the provision of very explicit instructions as may be required in a regulatory protocol or standard reference method.*

These reports are intended to provide guidance and to facilitate the use of consistent, appropriate, and comprehensive procedures for obtaining data on the toxicity to terrestrial or aquatic life of specific test substances or materials destined for or within the environment. Depending on the biological test method(s) chosen and the environmental compartment of concern, substances or materials to be tested for toxicity could include samples of chemical or chemical product, soil or similar particulate material, sediment or similar particulate material, effluent, elutriate, leachate, or receiving water. This report is to serve as a companion document to the biological test methods, in the Environmental Protection Series that describe toxicity tests with whole soil or similar particulate material. The methods described within this guidance manual for the collection, handling, transportation, storage, preparation and manipulation of whole soil are intrinsic to the acceptability and success of the recommended test methods involving soil toxicity evaluation. Although considerable guidance is provided within, key original references should be consulted for details.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Chemical Formulae

AC	Agriculture Canada
AAFC	Agriculture and Agri-Food Canada
AE	Alberta Environment
Ah	A horizon (surface mineral soil)
Al	Aluminium
ANOVA	Analysis of variance
ASTM	American Society for Testing and Materials
B	Boron
BR	Basal Respiration
BTEX	Benzene, toluene, ethylbenzene, and xylene
C	Carbon
°C	Degrees Celsius
Ca	Calcium
$CaCO_3$	Calcium carbonate
CanSIS	Canadian Soil Inventory System
CCME	Canadian Council of Ministers of the Environment
CEC	Cation exchange capacity
CLPP	Community level physiological profile
Cl^-	Chloride
cm	Centimetre(s)
COC	Chain-of-custody
Cu	Copper
Cu^{2+}	Copper (ion)
CV	Coefficient of variation
D	Minimum relative detectable difference/CV
DIN	Deutsches Institut für Normung (German Standards Institute)
DNA	Deoxyribonucleic acid
DQO(s)	Data quality objective(s)
EC	Electrical conductivity
Fe	Iron
DGGE	Denaturing gradient gel electrophoresis
GPS	Global positioning system
H	Hydrogen
HCl	Hydrochloric acid
H_2SO_4	Sulphuric acid
HDPE	High density polyethylene
IC25	Inhibiting concentration for a 25% effect
IC50	Inhibiting concentration for a 50% effect
ISO	International Organisation for Standardization
IUPAC	International Union of Pure and Applied Chemistry
K	Potassium
kg	Kilogram(s)
kGray	KiloGray
L	Litre(s)
LC50	Median Lethal Concentration
m	Metre(s)
m^2	Square metre(s)
MAL	Maximum acceptable limit
MDL	Minimum detection limit

MENV	Ministère de l'environnement du Québec (Quebec Ministry of the Environment)
Mg	Magnesium
Mn	Manganese
Mrad	Mega rad (1 million rads)
n	Number of samples
N	Nitrogen
Na	Sodium
Ni	Nickel
NH^{4+}	Ammonium
NO_3^-	Nitrate
NO_2^-	Nitrite
NRC	National Research Council
O	Oxygen
OECD	Organisation for Economic Co-operation and Development
P	Phosphorus
PAH	Polycyclic aromatic hydrocarbon
PCR	Polymerase chain reaction
PERD	Program of Energy Research and Development
PHCs	Petroleum hydrocarbons
POL	Program at Objective Level
PPE	Personal protective equipment
QA	Quality assurance
QC	Quality control
RNA	Ribonucleic acid
S	Sulphur
SAR	Sodium absorption ratio
SCC	Standards Council of Canada
S.I.	Le Système International d'Unités (International System of Units)
SIR	Substrate induced respiration
SOP	Standard operating procedure(s)
SRS	Simple random sampling
SSD	Species sensitivity distribution
StRS	Stratified random sampling
TDG	Transport of dangerous goods
TME	Terrestrial model ecosystem
TOC	Total organic carbon
USEPA	United States Environmental Protection Agency
UTM	Universal Transverse Mercator
µS	Microsiemens
VOC(s)	Volatile organic compound(s)
Z_α	Z statistic for Type I error probability
Z_β	Z statistic for Type II error probability
Zn	Zinc
%	Percentage or percent
°	Degree(s)
-	Minus
±	Plus or minus
>	Greater than
≥	Greater than or equal to
<	Less than
≤	Less than or equal to

Terminology

Note: the words defined in this section are italicized when first used in the body of the report according to the definition. All definitions are given in the context of the procedures in this report, and might not be appropriate in another context.

Grammatical Terms

Must is used to express an absolute requirement.

Should is used to state that the specific condition is recommended and ought to be met if possible.

May is used to mean "is (are) allowed to."

Can is used to mean "is (are) able to."

Might is used to express the possibility that something could exist or happen.

General Technical Terms

Active layer (soil) is the upper soil layer of permafrost which exists above the perennially frozen soil layer. The active layer normally thaws annually; its thickness depends on soil texture, water content, and peat cover. Most of the biological activity occurs in this layer (Danke, 1981).

A horizon (soil) is the mineral horizon at or near the surface of the soil. It is in the zone of leaching of materials in solution or suspension, or in the zone of maximum *in situ* accumulation of organic matter, or both (AAFC, 1998). Since it is a mineral soil layer, the A horizon contains 17% or less organic carbon (about 30% organic matter) by weight.

Ah horizon (soil) is a darkly coloured A horizon enriched in organic matter containing less than 17% organic carbon by weight (Bélanger and Van Rees, 2008).

A posteriori that (knowledge, decision) which is related to or derived by reasoning from observed facts.

A priori that (knowledge, decision) which is formed or conceived beforehand.

Artifact is an undesirable, detectable feature (e.g., chemical or physical change) of a substrate that has resulted from the activity or manipulations of those substrates.

Artificial soil is a laboratory-formulated soil, prepared to simulate a natural soil using a specific ratio of natural constituents of sand, clay, and peat. In certain instances, *artificial soil* may be used as a *negative control soil*, and as a diluent to prepare multiple concentrations of *site soil(s)* or chemical-spiked soil(s).

Assessment studies are projects that undertake the systematic gathering of information for the purpose of identifying and describing a specific condition in an ecosystem or environment.

Auger is a soil sampling device that removes soil from the ground by means of a rotating helical flighting where the material moves along the axis of rotation.

B horizon (soil) is the mineral horizon characterized by the enrichment with organic matter, oxides (e.g., aluminum oxide), or clay, or by the development of soil structure, or by a change in colour noting hydrolysis, reduction, or oxidation (AAFC, 1998). It is usually situated underneath the A horizon. Since it is a mineral soil layer, the B horizon contains 17% or less organic carbon (about 30% organic matter) by weight.

Battery of toxicity tests see *Test battery*

Bioavailability is a measure of the degree to which chemicals present in soil may be absorbed or metabolized by human or ecological receptors or are available for interaction with biological systems (ISO, 2005a).

Bulk density is the ratio of the mass of quantity of material (e.g., oven-dried soil) and the total volume occupied by this material (soil volume) (Culley, 1993). The soil volume is determined prior to drying. Note that bulk density is typically a volumetric mass, but it is commonly referred to as density (modified from ISO, 2005a).

Bulk soil samples are unconsolidated, typically large (> 1 L) point samples that consist of more than one individual block of soil removed from one sample location by a sampling device, and therefore are *point samples*, not *composite samples* (see *point* and *composite samples*). Bulk soil samples are often collected to satisfy the large volume requirements for biological testing.

C horizon (soil) is a mineral horizon comparatively unaffected by soil forming processes operating in the A and B horizons, except for the process of gleying (a process in which the oxygen supply in the soil profile is restricted due to soil moisture at saturation) and the accumulation of calcium and magnesium carbonates and soluble salts (AAFC, 1998). The C horizon is usually positioned below the B horizon. Since it is a mineral soil layer, the C horizon contains 17% or less organic carbon (about 30% organic matter) by weight.

Cation exchange capacity is the sum total of exchangeable cations that a soil can adsorb. It is sometimes called total-exchange capacity, base-exchange capacity or cation-adsorption capacity. It is expressed in milliequivalents per 100 grams of soil (or other adsorbing material such as clay) (AAFC, 1998).

Chain-of-custody is the documentation that establishes the control of a sample between the time it is collected and the time it is tested or analyzed. It applies to legal and non-legal samples to demonstrate that there was no tampering with or contamination of the sample during this time.

Clean soil is a soil that contains no *contaminant* at concentrations that cause discernable distress to organisms or reduce their survival, growth, or reproduction.

Composite sample(s) are soil samples consisting of point or bulk samples combined from two or more sample locations at a site (Crepin and Johnson, 1993)

Concentration means, for biological tests, the ratio of weight of test substance or material (e.g., soil contaminant) to the weight of soil and is frequently expressed as the weight of test substance or material per kg of dry soil (mg/kg). Concentration might also be expressed as a percentage of the test substance or material (e.g., contaminated site soil per dry weight of soil).

Conductivity is a numerical expression of the ability of an aqueous solution to carry an electric current. This ability depends on the concentration of ions in a solution, their valence and mobility, and on the solution's temperature. Conductivity is reported as decisiemens per metre (dS/m) or as micromhos per centimetre (μmhos/cm); 1 dS/m = 1000 μmhos/cm.

Consolidated soil sample (see also *unconsolidated soil sample*) is synonymous with undisturbed sample and is a sample obtained from soil using a method designed to preserve the soil structure (ISO, 2005a).

Contaminant is a substance or material that is present in a natural system, or present at increased concentrations, often because of some direct or indirect human activity. The term is frequently applied to substances or materials that are present at concentrations that have the potential to cause adverse biological effects. The Canadian Council of Ministers of the Environment (CCME) defines contaminant as "any substance present in an environmental medium at concentrations in excess of natural background" (CCME, 2006).

Contaminated soil is soil containing chemical substances or materials (see *contaminant*) at concentrations that pose a known or potential threat to environmental or human health.

Core sample is a sample of soil that has been collected using a *corer*.

Corer (e.g., core, ring, or cylinder sampler) is a device used to collect a column of soil (e.g., a *core sample*) which represents the vertical distribution of the physical and chemical characteristics of the soil.

Data quality objectives (DQOs) are pre-defined criteria for the quality of data generated or used in a particular study so as to ensure that the data are of acceptable quality to meet the needs for which they were collected.

De-ionized water is water that has been purified by passing it through resin columns or a reverse osmosis system, for the purpose of removing ions such as calcium or magnesium.

Ecological risk assessment is the process of risk analyses and evaluation of the adverse effects of contaminated environmental media (e.g., air, soil, water) on non-human organisms with respect to the nature, extent, and probability of the occurrence of these effects (ISO, 2005a).

Ecotoxicology is a branch of science that studies the integrated ecological and toxicological effects of chemical pollutants on populations, communities and ecosystems with the fate (transport, transformation, and breakdown) of such pollutants in the environment (modified from Forbes and Forbes, 1994).

Ecozone is an area in which organisms and their physical environment endure as a system and are categorized based on broad, naturally occurring common characteristics such as landforms, soils, water features, vegetation, and climate (EC, 2005b).

Environmental risk assessment is the process of risk analyses and evaluation of the adverse effects of contaminated environmental media (e.g., air, soil, water) on both humans and non-human organisms with respect to the nature, extent, and probability of the occurrence of these effects (ISO, 2005a).

F horizon (soil) is an organic horizon at the surface of mineral soil developed primarily from the accumulation of leaves, twigs, and woody materials. The F horizon is characterized by an accumulation

of partially decomposed organic matter in which the original structures are difficult to recognize (AAFC, 1998). The F horizon is positioned below the L horizon and above the H horizon. The H horizon is positioned above the top mineral (A) horizon. Since it is an organic soil layer, the F horizon contains greater than 17% organic carbon (about 30% organic matter) by weight.

Fertility (soil) refers to the potential of a soil to supply nutrients in the amounts, forms, and proportions required for plant growth. Soil fertility is measured directly in terms of the ions and compounds important for plant nutrition. The fundamental components of fertility are the essential nutrients (macronutrients including C, H, O, N, P, K, Ca, Mg, S and micronutrients including Fe, Mn, Mo, B, Cu, Zn, and Cl).

Guild is a grouping of organisms, especially animal species that occupies a common niche in a given community, characterized by the exploitation of environmental resources (e.g., food, microhabitat, etc.) in the same way.

H horizon (soil) is an organic horizon at the surface of mineral soil developed primarily from the accumulation of leaves, twigs, and woody materials. The H horizon is characterized by an accumulation of decomposed organic matter in which the original structures are not discernable. This horizon differs from the F by having greater humification due to the action of organisms (AAFC, 1998). The H horizon is positioned below the F horizon and is frequently intermixed with mineral grains, especially near the junction with the mineral horizon. The H horizon is positioned above the top mineral (A) horizon. Since it is an organic soil layer, the H horizon contains greater than 17% organic carbon (about 30% organic matter) by weight.

Isopleth is a line joining identical soil types or contaminate concentrations of equal value on a map.

L horizon (soil) is an organic horizon at the surface of mineral soil developed primarily from the accumulation of leaves, twigs, and woody materials. The L horizon is characterized by a layer of organic matter where the original organic material structures are easily discernable (AAFC, 1998). LFH horizons often occur in combination, and L is the top layer, followed by F then H. The H horizon is positioned above the topmost mineral (A) horizon. Since it is an organic soil layer, the L horizon contains greater than 17% organic carbon (about 30% organic matter) by weight.

Legal sample is a sample that is collected with a view to prosecution (e.g., the toxicity results might be admissible in court). A legal sample is considered to be representative of the material being sampled and must be free of contamination by foreign substances during or after sampling. The origin of the sample, time and method of collection must be identified, and the chain-of-custody clearly documented. Legal samples are transported in labeled containers with a seal, stored in a secure and locked place, and processed as soon as possible after collection.

Macrofauna are soil-dwelling invertebrates with a body width > 2 mm but < 20 mm such as annelids (e.g., earthworms), molluscs (e.g., snails), isopods (e.g., pillbugs), chilopods (e.g., centipedes), and diplopods (e.g., millipedes) (Swift *et al.*, 1979).

Material is the substance or substances from which something is made. A material has more or less uniform characteristics. Soil, sediment, or surface water are materials. Usually the material contains several or many substances.

Mesofauna are small soil-dwelling invertebrates with a body width > 100 μm but < 2mm such as acari (e.g., mites), collembola (springtails), and enchytraeids (potworms) (Swift *et al.*, 1979).

Microbial functional tests are observations of system attributes or a survey of the microbial system. They include: determination of biomass, estimate of number of culturable individuals, and community composition or activity.

Microfauna are small, often microscopic soil-dwelling organisms with a body width < 100 μm such as protozoa, bacteria, nematoda, and fungi (Swift *et al.*, 1979).

Mineral soil consists predominantly of, and having properties determined by, mineral matter. Mineral soil consists of less than 17% organic carbon (less than about 30% organic matter) by weight but may contain organic surface layers (AAFC, 1998).

Moisture content is the percentage of water in a sample of test soil, based on its wet or dry mass. It is determined by measuring both the wet and dry weights of a subsample of the soil and subtracting the dry weight from the wet weight and dividing by either the dry weight or the wet weight.

Negative control soil is soil that does not contain concentrations of one or more contaminants which could affect the survival, reproduction, growth or behaviour of the test organisms. *Negative control soil* might be natural soil from an uncontaminated site, or artificial (formulated) soil. This soil must contain no added test material or substance, and must enable acceptable survival and performance of the test organisms during the test. The use of *negative control soil* provides a basis for interpreting data derived from toxicity tests using *test soil(s)*.

O horizon (soil) occurs in organic and mineral soils and is developed mainly from mosses, plant debris, and woody materials. It can occur at any depth in organic soil and at the surface of mineral soils to a depth of less than 40 cm and contains greater than 17% organic carbon (about 30% organic matter) by weight (AAFC, 1998).

Organic matter (OM) in soil consists primarily of plant and animal residues, at different stages of decomposition, including soil humus. The accumulation of OM within soil is a ratio between the return or addition of plant and animal residues and their subsequent loss due to the decay of these residues by soil organisms. For many types of soils, the following equation (AESA, 2001) is suitable for estimating the total OM content of soil from *total organic carbon (TOC)* measurements: %OM = %TOC × 1.78 (See also *total organic carbon*).

Organic soil is soil that is comprised largely of organic materials. Organic soils contain greater than 17% organic carbon (about 30% organic matter) by weight and include most of the soils commonly known as peat, muck, bog, or fen soils. Most organic soils are saturated with water for prolonged periods because of their location in poorly drained areas and humid climates; however, there are some upland organic soils of forest origin that are well to imperfectly drained (AAFC, 1998).

Permafrost is a thermal condition of soil or rock in which temperatures below 0°C persist over at least two consecutive winters and the intervening summer (Danke, 1981).

pH is the negative logarithm of the activity of hydrogen ions in gram equivalents per litre. The pH value expresses the degree or intensity of both acidic and alkaline reactions on a scale from 0 to 14, with 7

representing neutrality; numbers less than 7 indicate increasingly greater acidic reactions and numbers greater than 7 indicate increasingly basic or alkaline reactions.

Point sample(s) are individual blocks of soil removed from one sample location by a sampling device (e.g., a soil core).

Quality assurance (QA) refers to the management and technical practices (e.g., planning, control, assessment, reporting, remedial action) designed to ensure an end product of known or reliable quality.

Quality control (QC) refers to the techniques and procedures used to measure and assess data quality and the remedial actions to be taken when data quality objectives are not realized.

Receiving container is a container that is used to hold field-collected samples for handling on-site. It may or may not be the sample container.

Redox potential (also known as the oxidation-reduction potential) is a measure (in volts) of the affinity of a substance for electrons relative to hydrogen.

Reference soil is typically clean field-collected soil that is selected for use in a particular study or toxicity test with one or more samples of test soil. Reference soil used in a test frequently exhibits physical and chemical properties (e.g., texture, total organic carbon, pH, etc.) closely matching those of the test soil sample(s) except that it is free of the contamination being assessed. Reference soil is used to describe matrix effects in the test, and may also be used as a diluent to prepare a series of exposure concentrations of the test soil.

Reference site is the uncontaminated site from which reference soil samples are collected. Reference sites could be adjacent or in the vicinity of the study site or from the same area (e.g., city or region) as the study site.

Remediation is the removal of contaminants from soil, sediment, surface or groundwater (or removal of one or more of the contaminated media itself) through the use of physical, chemical, or biological processes for the protection of human health and/or the environment.

Risk is the probability or likelihood that an adverse effect will occur.

Risk management is the management of a contaminated site to prevent, minimize, or mitigate damage to human health or the environment. Risk management can include both direct physical actions (e.g., removal, destruction, and containment of toxic substances) and institutional controls (e.g., zoning designations or orders).

Salinity is a term used to describe the salt content in a material. A saline soil is a non-sodic soil containing sufficient soluble salts to impair its productivity. The conductivity of the saturation extract of a saline soil is greater than 4 dS/m (at 25°C) and the pH is usually less than 8.5 (AE, 2001).

Sample(s) is a portion of material (e.g., soil) selected from a larger quantity of material.

Sample container is a container into which a field-collected sample is placed directly from the sampler and can be, but is not necessarily, the final receptacle for soil to be shipped to the laboratory. For example, soil could be placed into a large bin or bucket (both *sample containers*) for compositing with other subsamples prior to shipping to a laboratory.

Sampling design see *sampling strategy*

Sample location is the place where sampling occurs within a study site (and is considered the same as a sample point) (ISO, 2005b; IUPAC, 2005).

Sample size is the actual volume (L or m^3), weight (g), or dimensions (diameter and length) of a sample of soil.

Sample volume is the volume (L, m^3) of a sample.

Sampler refers to the device used to collect samples or subsamples.

Sampling is the act of collecting samples.

Sampling depth is the depth of soil from which the sample is collected.

Sampling strategy is a process by which the type, location and collection method of samples is determined and is considered the same as a *sampling design* (ISO, 2005b; IUPAC, 2005).

Sieving is the manual pressing of soil particles through a sieve of a particular mesh size.

Site means a delineated tract of land that is being used or considered as a study site, usually from the perspective of it being contaminated or potentially contaminated.

Site soil is a field-collected sample of soil, taken from a site either contaminated or potentially contaminated with one or more chemicals, and intended for use in biological tests.

Soil is the upper layer of the Earth's crust transformed by the physical, chemical, and biological weathering of rocks and the decomposition and recycling of nutrients from organic matter originating from plant and animal life. Its physical and chemical characteristics are influenced by microbial and invertebrate activities therein, and by anthropogenic activities. Soil is the whole, intact material representative of the terrestrial environment that has had minimal manipulation following collection or formulation.

Soil horizon is a layer of mineral or organic soil material approximately parallel to the land surface that has characteristics altered by processes of soil formation. It differs from adjacent horizons in properties such as color, structure, texture, and consistence and in chemical, biological, or mineralogical composition. The major mineral horizons are A, B, and C. The major organic horizons are L, F, H, and O (AAFC, 1998).

Soil properties are the inherent physical and chemical characteristics of soil such as pH, texture, major ion concentrations, cation exchange capacity, redox potential, salinity, total organic carbon, etc.

Soil sample is a portion of soil material selected from a larger quantity of material.

Study area refers to a locale that contains the study site to be assessed, as well as adjacent landscape that might influence the conditions of the study site. *Study area* is often referred to as "adjacent area" in contaminated site assessments.

Study site refers to a smaller location within the study areas that is to be assessed, monitored, or from which the samples are to be collected. *Study site* is often referred to simply as "site" or "the Site" in contaminated site assessments.

Substance is a particular kind of material having more or less uniform properties. The word substance has a narrower scope than material and might refer to specific chemical or chemical products.

Subsample is a representative part of a sample that is studied in order to gain information about the characteristics or infer properties of the sample. A subsample must represent a single time of collection. If collected over time, the observations would fall in the category of repeated measures.

Subsampling refers to the act of collecting subsamples.

Test soil is a sample of field-collected soil or formulated and/or field-collected chemical-amended soil to be evaluated for toxicity to biological organisms.

Test battery is a combination of several toxicity tests, normally using different species of test organisms (e.g., a series of soil toxicity tests using earthworms, plants, arthropods, or a series of soil toxicity tests using several species of plants), different biological endpoints (e.g., lethal and various sublethal), different durations of exposure (e.g., acute and chronic) and different types of tests (e.g., single-species toxicity tests and functional tests).

Texture (soil) is defined based on a measurement of the percentage, by weight, of sand, silt, and clay in the mineral fraction of soils. Classification as to texture confers information on the general character and behaviour of substances in soils, especially when coupled with information on the structural state and organic matter content of the soil. Texture in the context of this guidance document is described according to the Canadian System of Soil Classification (AAFC, 1998) not the Unified Soil Classification, the United States Soil Conservation Service Classification or any other soil classification system used for soil science, engineering or geology. Soil texture is determined in the laboratory by measuring the particle-size distribution using a two-step procedure whereby the sand particles (coarse fragments) are initially separated by sieving from the silt and clay particles; followed by separation of the silt and clay particles by their sedimentation in water. Textural classification is based on specific ranges in relative quantities of sand (0.05 to 2.0 mm diameter), silt (0.002 to 0.05 mm diameter) and clay ($\leq$ 0.0002 to $\leq$ 0.002 mm diameter) (AAFC, 1998). There are three main textural classes:

i. coarse texture (sands, loamy sands, sandy loams);

ii. medium texture (loams, silt loams, silts, very fine sandy loams); and,

iii. fine texture (clays, silty clay loams, sandy clay loams, silty clays and sandy clays).

Further distinction as to texture (e.g., "sandy clay," "silt loam," "loam") can be made based on classification schemes using the relative amounts (percentages) of sand, silt, and clay in the soil (AAFC, 1998).

Total organic carbon (TOC) refers to the organic carbon content of soil exclusive of carbon from undecayed plant and animal residues, as determined by dry combustion analysis (Tiessen and Moir, 1993). See also *organic matter*.

Unconsolidated soil sample (see also *consolidated soil sample*) is synonymous with disturbed sample and is a sample obtained from soil without any attempt to preserve the soil structure (ISO, 2005a).

Water-holding capacity (WHC) refers to the maximum quantity of water that a soil can retain, following complete saturation. It is usually determined gravimetrically, and is generally expressed as the percentage of water (by mass; water weight: dry soil weight) retained in a sample of soil that has been saturated with water.

Wetlands are areas in which soils are water-saturated for a sufficient length of time such that excess water and resulting low soil oxygen levels are principal determinants of vegetation and soil development. Wetlands will have a relative abundance of hydrophytes in the vegetation community and/or soils featuring "hydric characteristics" and have both semi-aquatic and semi-terrestrial characteristics [(Mackenzie and Moran, 2004) as cited in UMA, 2008].

Toxicological and Statistical Terms

Acute means within a short period of exposure (seconds, minutes, hours, or a few days) in relation to the lifespan of the test organism.

Acute toxicity is a discernable effect (lethal or sublethal) induced in the test organisms within a short period (usually a few days) of exposure to test soils.

Biological Test is any test in which the strength or potency of a substance is measured by the response of living organisms. This may include single species tests, multiple species tests, or microbial tests. *Toxicity test* is more specific to single and/or multiple species tests only.

Chronic means occurring within a relatively long period of exposure (weeks, months, years), usually a significant portion of the lifespan of the organism (e.g., 10% or more).

Chronic toxicity implies adverse effects during or after relatively long-term exposure to one or more contaminants, which are related to changes in reproduction, growth, metabolism, ability to survive, or other biological variables (e.g., behaviour) being observed.

Coefficient of variation (CV) is the standard deviation (SD) of a set of data divided by the mean of the data set, expressed as a percentage. It is calculated according to the following formula: CV (%) = 100 (SD/mean).

Cosine rule is a trigonomic relationship that allows any unknown length or angle of a triangle to be calculated provided that at least some of the lengths and angles are known. The *cosine rule* relates the lengths of the triangle sides to the cosine of one of its angles and is expressed mathematically as: $a^2 = b^2 + c^2 - 2bc[\cos(A)]$. Note that a, b, and c are lengths and A, B, and C are angles, of the triangle.

Definitive (soil toxicity test) means decisive (as opposed to preliminary, *screening* test). See also *screening test*. This term is generally used to describe terrestrial plant toxicity tests with longer exposure durations relative to *screening* tests.

ECx is the concentration (e.g., % or mg/kg) of substance(s) or material(s) in soil that is estimated to cause some defined toxic effect on x% of the test organisms. In most instances, the ECx and its 95% confidence intervals (95% CI) are statistically derived by analyzing the percentages of organisms affected (e.g., showing an avoidance response) at various test concentrations, after a fixed period of exposure. The duration of exposure must be specified (e.g., 48 h or 72 h). The ECx describes sublethal quantal effects (e.g., effects with binomial responses such as avoidance or no avoidance) and is not applicable to continuous quantitative effects (e.g., effects that can be measured along a numerical continuum such as length or weight) (see *ICp*). In most instances, the EC50 is calculated, although depending on the study objectives, the EC20 or EC25 may also be calculated.

Endpoint means the measurement(s) or value(s) that characterize the results of a test (e.g., IC25). It also means the response of the test organisms that is measured (e.g., death, number of progeny produced, amount of nitrate extracted from soil).

Environmental toxicology is a branch of toxicology with the same general definition; however, the focus is on ecosystems, natural communities, and wild living species, without excluding humans as part of the ecosystems.

Functional test is a determination of the effect of a substance or material (e.g., contaminants in soil) on biologically mediated processes in soil such as nutrient cycling, organic matter breakdown, or soil respiration under defined conditions. Measurements usually include quantification of the increase or decrease of substances produced or consumed as a result of these processes such as nitrate and carbon dioxide production, or the consumption of organic matter, etc.

Geostatistics are statistics used to describe the spatial dependence of natural properties of soil, soil contamination, or soil toxicity. Soil properties, contamination, and toxicity vary continuously in space; the values at locations close together are more similar than those farther apart and this spatial dependence can be described by the use of geostatistics.

ICp is the inhibiting concentration for a specified percent *(p)* effect. It represents a point estimate of the concentration of test substance or material that causes a designated percent inhibition *(p)* compared to the control in a quantitative (continuous) biological measurement such as number of progeny produced or size attained by individuals at the end of a test.

LCx is the concentration (e.g., % or mg/kg) of substance(s) or material(s) in soil that is estimated to be lethal to x% of the test organisms. In most instances, the LCx and its 95% CI are statistically derived by analyzing the percent mortalities in five or more test concentrations, after a fixed period of exposure. The duration of exposure must be specified (e.g., 7-day LCx or 14-day LCx). In most instances, the LC50 is calculated, though depending on the study objectives the LC20 or LC25 may also be calculated.

Lethal means causing death by direct action. Death of test organisms is defined as the cessation of all visible signs of movement or other activity indicating life.

Replicate samples are samples of soil collected independently from the same sampling location. A single soil sample from a sampling location is treated as one replicate. Additional samples are considered to be additional replicate soil samples if they are treated identically (regardless if they are point or composite samples from the same location) but stored in separate sample containers (e.g., not composited or, if already composite samples, not composited further).

Replicate treatment. Test chamber, test unit, or experimental unit refers to a single test chamber in a laboratory containing a prescribed number of organisms in either one treatment of the test material or substance, or in the reference treatment(s). A replicate of a treatment must be an independent test unit; therefore, any transfer of organisms or test material from one test unit to another would invalidate statistical analyses based on the replication. If individual replicate treatments are derived from the same soil field-collected sample, then the soils in the replicate treatments are in fact sub-samples of the field samples.

Screening (soil toxicity test) means a preliminary soil toxicity test, performed to provide an initial indication of the toxicity of the test substances or material under defined conditions and/or to assist in choosing the range of concentrations to be used in a definitive multi-concentration test (i.e., range-finding).

Species sensitivity distribution (SSD) is constructed from site-specific data to determine the probability distribution of some measure of toxicity caused by contaminated site soil (or one or more specific chemicals) in a population of plant and animal species (i.e., in the battery of test organisms used to assess the site soil).

Sublethal (toxicity) means detrimental to the organism, but below the concentration or level of contamination that directly causes death within the test period.

Sublethal effect is an adverse effect on an organism, below the concentration or level of contamination that directly causes death within the test period.

Toxic means poisonous. A toxic chemical or material can cause adverse effects on living organisms, if exposure pathways are complete and the concentration sufficient. Toxic is an adjective or adverb, and should not be used as a noun; whereas toxicant is a legitimate noun.

Toxicant is a toxic substance or material.

Toxicity is the inherent potential or capacity of a substance or material to cause adverse effect(s) to living organisms. These effect(s) could be lethal or sublethal.

Toxicity test is a determination of the effect of a substance or material to a group of selected organisms under defined conditions. A toxicity test involving samples of test soil usually measures: a) the proportion of organisms affected; and/or, b) the response after exposure of the test organisms to the whole sample (e.g., undiluted site soil) or some concentration thereof.

Toxicology is a branch of science that studies the toxicity of substances, materials, or conditions. There is no limitation on the use of various scientific disciplines, field or laboratory tools, or studies at various levels of organization (i.e., molecular, single species, populations, or communities). Applied toxicology would normally have a goal of defining the limits of safety of chemical or other agents.

Treatment refers to a specific test soil (e.g., a site soil, negative control, or reference soil) from a particular sampling location, or a concentration of chemical-amended soil or a mixture of test soil diluted with clean soil prepared in the laboratory. Test soils representing a particular treatment are typically replicated in a toxicity test.

Variogram is a statistical measure of spatial variation of a variable (ISO, 2005a).

Acknowledgements

This document was written by Natalie Feisthauer (Feisthauer Environmental, Cambridge, Ontario) with the direct technical input from Barry Zajdlik (Zajdlik & Associates Inc., Rockwood, Ontario), Kathryn Bessie and Natasha Harckham (EBA Engineering Consultants Ltd., Calgary, Alberta), and Jessica Velicogna (Environment Canada), who is sincerely thanked for also collating and incorporating responses to reviewers' comments for the final draft of the report. Richard Scroggins (Biological Assessment and Standardization Section, Environment Canada, Ottawa, Ontario) was the Scientific Authority for the project and provided technical input and direction throughout the work. Jörg Römbke (ECT Oekotoxikologie GmbH, Flörsheim, Germany) and Gladys Stephenson (Stantec Consulting Ltd., Guelph, Ontario) provided expert advice and scientific guidance as members of the Technical Advisory Group. The information in Appendix C (Case Studies for Sampling Strategy Selection and Calculating the Number of Samples) was provided by Barry Zajdlik, information in Appendices B and D (Ecological Classification Resources and Geostatistics, respectively) was provided by Kathryn Bessie and Natasha Harckham, additional information in Appendices B and E (Ecological Classification Resources and Useful Web sites, respectively) was provided by Jessica Velicogna, and information in Sections 6.2 and 6.5 (Shallow/Stony Soils and Wetland Soils, respectively) was provided by Jessica Velicogna with direct technical input from Doug Bright (AECOM Canada, Victoria, British Columbia). We are grateful for permission granted to describe the information provided in Case Studies #1, #2, and #3 presented in Appendix A. Photographs of most of the Canadian Soil Order profiles provided in Appendix E (Soil Classification) have been reproduced with permission of the Soil Landscapes of Canada Working Group (2007); the two photographs of Vertisolic soils have been reproduced courtesy of the NRC Research Press (AAFC, 1998). Maps depicting the geographic distribution of the Soil Orders in Canada presented in Appendix E were reproduced with permission of the University of Saskatchewan (2009).

We gratefully acknowledge the invaluable assistance provided by members of the Environment Canada Scientific Advisory Group (SAG) who provided many useful comments, suggestions, and materials (photographs, references, documents, etc.) on drafts at one or more stages of development: Doug Bright, Kathryn Bessie, Mark Crane (Watts & Crane Associates, Faringdon, UK), Gordon Dinwoodie (Alberta Environment, Edmonton, Alberta), Roman Kuperman (US Army, Maryland, USA), John Jensen (National Environmental Research Institute, Silkeborg, Denmark), Juliska Princz (Environment Canada, Ottawa, Ontario), and Agnès Renoux (Sanexen Environmental Services, Varennes, Québec).

In addition to the SAG members who reviewed this document, the following people also reviewed the initial and/or final drafts and provided useful comments: Stephan Jänsch (ECT Oekotoxikologie GmbH, Flörsheim, Germany) and Barry Zajdlik.

Many thanks to the following people for providing diagrams and photographs of soil sampling and sampling devices: Kathryn Bessie, Doug Bright, Natasha Harckham, Juliska Princz, Agnès Renoux, Jörg Römbke, Ben Smith (Stantec Consulting Ltd, Guelph, Ontario), and Gladys Stephenson.

Introduction

1.1 Background

Soil is the uppermost stratum of the Earth's crust that has been transformed by weathering and biological processes to comprise a complex and heterogeneous mixture of mineral particles, organic *material*, water, air, and living organisms. Protection of soil as a valuable resource is a priority in Canada.

Important causes of soil degradation include erosion, loss of *organic matter*, and chemical contamination. Soil in many areas of Canada has become degraded through contamination with organic and inorganic *contaminants* as a result of anthropogenic activities such as: resource extraction and processing, manufacturing, agriculture, and urbanization.

The Contaminated Sites Management Working Group, a federal interdepartmental committee, established in 1995, assigned a high priority to the standardized assessment and *remediation* of contaminated land across Canada. The Canadian Council of Ministers of the Environment (CCME) has published guidance on the assessment of contaminated land (CCME, 1993a, b) and guidance for conducting multi-tiered *ecological risk assessments* on contaminated land (CCME, 1996a; 1997). To date, most ecological risk assessments of contaminated land have relied solely upon a chemical-specific approach in which contaminant *concentrations* in soil or soil pore water have been used to estimate *toxicity*. Increasingly, however, *biological tests* that directly measure the toxicity of *contaminated soils* are being integrated into contaminated land risk assessments and are significantly contributing to more effective contaminated land management (Stephenson *et al.*, 2002; Lanno, 2003; Thompson *et al.*, 2005).

It is recognized that methods used to *sample*, handle, store, transport, prepare, and manipulate contaminated soils can significantly influence the results of chemical analyses. This is also true for the results of biological tests with contaminated soils. While extensive guidance exists for the *sampling*, collection, handling, and preparation of soils for chemical analyses, no such guidance exists for biological testing.

1.2 Purpose, Scope, and Application of Guidance

The use of terrestrial *toxicity testing* with species directly exposed to contaminants in soil (e.g., plants, macro- and micro-invertebrates, fungi, bacteria) can provide valuable information to support decision-making processes for the management of contaminated *sites* in Canada. Site-specific terrestrial toxicity testing results can be: one line of evidence in a weight-of-evidence approach used to manage contaminated land; used as a tool to augment site data; or, used to investigate the mechanistic interaction between contaminants and the ecological receptors (e.g., plants, animals, and microorganisms) that inhabit the site.

Toxicity data are used to assess risk associated with contaminants in soil to soil-dwelling organisms. Risk to organisms through food-web (i.e., trophic) transfer of soil contaminants can also be assessed through bioaccumulation studies with soil-dwelling organisms. In addition, data generated by site-specific testing can be used for the derivation of site-specific remediation objectives based on direct contact of contaminants with soil-dependent organisms. Environment Canada has recently published standardized biological test methods for earthworms, plants, and collembola (soil arthropods) (EC, 2004a, 2005a, 2007a, respectively) tailored to provide guidance for conducting toxicity tests of contaminated soil on a site-specific basis.

Although some guidance is now available for conducting site-specific toxicity tests (e.g., EC, 2004a; 2005a, 2007a), what is currently lacking is comprehensive guidance for the sampling,

handling, transportation, storage, preparation, and manipulation of soil for biological testing. There is a wealth of guidance for the sampling and collection of soil for chemical characterization (Mason, 1992; CCME 1993a, b; ISO 2002a, b, c, 2003a, 2005b, c, d; USEPA 1986, 2002a, b, 2006). However, the objectives and procedures for sampling, collecting, and preparing soils for use in biological testing are often significantly different from those for soils collected for chemical analyses. The need for such comprehensive guidance was first identified as a priority by participants at a workshop on toxicity testing applied to soil ecotoxicity held by the National Research Council in 1995 (NRC, 1995). More recently, participants at a multi-disciplinary workshop organized by Environment Canada also agreed that development of guidance for soil collection and preparation for biological testing should be a high priority (EC, 2003a). The key objective of this workshop was to prioritize new toxicity test methodologies and guidance for assessing the impacts of mixtures of contaminants in soil using terrestrial species of ecological relevance to Canadian soil systems. The need for such guidance was also identified as a priority issue by a working group for the Soil and Groundwater Program at Objective Level (POL) convened by the Program of Energy Research and Development (PERD) (Environment Canada) in March, 2007.

Biological testing of soils has been used to support site-specific ecological risk assessment programs (e.g., Tier 3 risk assessments). Until recently there has been no regulatory requirement for the use of terrestrial toxicity data to support management decisions regarding contaminated sites in Canada. Alberta Environment (AE), however, published guidance for deriving site-specific Tier 2 soil and groundwater remediation guidelines (AE, 2007a). This guidance describes how to use terrestrial toxicity testing to develop site-specific objectives for direct exposure pathways (e.g., earthworms, plants, etc.) at a Tier 2 level. Specific guidance on how to utilize toxicity testing data in Tier 2 risk assessments is provided in a draft protocol that accompanied the AE guidelines (AE, 2007b); however, at present this approach is restricted to soils contaminated with benzene, toluene, ethylbenzene, and xylene (BTEX) and petroleum hydrocarbons.

1.2.1 Purpose of Guidance Document

The guidance provided herein is targeted for site and risk assessors and contaminated land managers. It includes guidance for methods and procedures for designing a sampling program, and for sampling, handling, transporting, storing, and preparing contaminated soil for biological testing. It is anticipated that site-specific biological testing of soils will usually be implemented as part of a higher level of risk assessment (e.g., Tier 2 or Tier 3). It should be noted that the guidance herein is provided under the assumption that some data on the characterization of the chemical and soil physical and chemical properties of the land under investigation are already available from earlier phases of the contaminated land management process.

1.2.2 Scope of Guidance

The scope of this guidance encompasses the sampling, handling, transportation, storage, and preparation of contaminated soil primarily for single-species toxicity and microbial testing; however, this guidance is also applicable for multi-species microcosm tests that could be applied in specific contaminated land assessment situations.

Data from toxicity tests should not be used as stand-alone evidence of the effects of soil contaminants at a site. Instead, toxicity test data should be used in conjunction with chemical analytical data, and, ideally, the results of ecological field surveys. Toxicity test data can be used as one line of evidence in a weight-of-evidence approach. This integrated approach using chemistry, toxicity, and ecological data is similar to the TRIAD approach used to assess and manage risk associated with contaminated sediment and is considered to provide more accurate information of the risk associated with soil contamination than using only one line of evidence since it decreases uncertainty (Jensen *et al.*, 2006a; Mesman *et al.*, 2006; EC, 2008). More detail on the use of multiple lines of evidence in ecological risk assessments and

contaminated land management can be found in the Environment Canada guidance document on the use of single-species tests in *environmental toxicology* (EC, 1999) and in Jensen and Mesman (2006).

When soils are contaminated with either very high or very low concentrations of a compound(s), it is often not necessary to use biological testing to support a contaminated site management decision. In these cases the use of biological tests does not reduce uncertainty, because they are already small.

The following is a list of tests and studies that are *not* within the scope of this guidance:

1. Characterization of soil contamination at a site.[1]
2. Characterization of soil physical and chemical properties at a site.
3. *In situ* field testing (e.g., earthworm exposure, litter bag assessment, bait lamina strips).[2]
4. Ecological surveys (e.g., plant, invertebrate, wildlife).[2]
5. Field sampling of plant tissue, invertebrates or wildlife.[2]
6. Field measurements of soil microbial structure (e.g., DNA probes).
7. Microbial degradation tests.
8. Invertebrate taxonomic identification in the field or of soil cores or *bulk soil samples*.
9. Soil leachate or elutriate collection and testing.
10. Soil pore water collection and testing.

[1] Toxicity testing can be used to help focus chemical sampling efforts in areas at a site where the bioavailability (i.e., toxicity) of contaminants has been confirmed by toxicity testing of previously collected soil samples.

[2] For an ecological risk assessment of contaminated soil, any sampling plan designed for biological testing should give statistical consideration to field tests or ecological surveys so that the toxicity test and field study results can be integrated and linked with soil chemistry data to provide a more precise estimate of adverse ecological effects (Linder *et al.*, 1992).

1.2.3 Benefits and Limitations of Biological Testing Approaches in Contaminated Land Assessment and Remediation

Currently the approach most frequently used to estimate the risk associated with contaminated soil is through the quantification of chemical concentrations of individual compounds or groups of compounds in *site soil*. If the concentrations of one or more compounds of concern exceed a generic regulatory criterion, standard or guideline for a specific compound or group of compounds, then the soil is considered to constitute a hazard to ecological (and/or human, depending on how the value was derived) receptors. Regulatory criteria, standards or guidelines are generally chemical-specific, and are derived using generic single-species toxicity data and usually incorporate an uncertainty, adjustment or safety factor *in lieu* of site-specific data on *bioavailability*, contaminant interaction etc.

The integration of measured concentrations of contaminants in soil with site-specific biological test data (e.g., single-species toxicity tests, *microbial functional tests*) provides a more comprehensive and accurate estimate of the site-specific risk of contaminated soil to ecological receptors in direct contact with soil. This is because data from site-specific biological testing is an integration of the inherent toxicity of the contaminant(s) and their mixtures in the soil with the bioavailability of the contaminant(s) under the specific conditions at the site. These site-specific conditions that modify contaminant bioavailability include, but are not limited to: 1) soil physical and chemical characteristics; 2) contaminant speciation and degree of aging and/or weathering; and, 3) the integrated toxicity of contaminant mixtures (e.g., additive, antagonistic, or synergistic) (Greene *et al.*, 1989). The use of site-specific biological testing also facilitates the generation of toxicity data for selected test species that are more ecologically relevant to the site under investigation than species whose data are used to derive guidelines. Biological testing also detects the effects of all *toxic* contaminants in the soil, including those not measured or detected by chemical analyses. The demonstrated benefits of using biological tests as part of contaminated site assessments

have been documented for over 20 years (Thomas *et al.*, 1986; Athey *et al.*, 1987; Spurgeon *et al.*, 2004; Dirven-van Breemen *et al.*, 2006; Römbke *et al.*, 2006a). There are a number of different testing approaches that can be used to assess the toxicity of soil samples collected from a contaminated site. The three principal approaches are:

1. Testing one or more soil samples from a contaminated site and comparing the results to a suitable reference or *negative control soil*.

2. Testing soil samples from a site that represent an exposure concentration series due to their different contamination levels or by collecting soil samples along a known chemical concentration gradient at the site.

3. Testing soil samples that represent an exposure concentration series by diluting contaminated site samples with a negative control or *reference soil*.

Two additional approaches that might be considered in special situations include:

4. When co-contaminant(s) are present in soil samples, testing site soil samples that have been amended with the co-contaminant(s) to a specified and equal concentration among all site samples.

5. When co-contaminant(s) are present in soil samples, testing a site soil sample that has been amended with the co-contaminant(s) to a specified concentration but that has been diluted with a reference or negative control soil to create an exposure concentration series.

Detailed discussions of these approaches are provided in Stephenson *et al.* (2008) and Stantec (2004).

Along with the positive attributes of using biological testing in contaminated land management, of course, there are limitations to the application and use of these tests and test data. A summary of both benefits and limitations of the use of biological test data is provided in Table 1.

In consideration of both the benefits and limitations of using biological tests in contaminated site assessments, it is strongly recommended that any testing program use a battery of standard tests. The value of using more than one test type and test species in an ecotoxicity evaluation cannot be overstated. The use of different types of tests and species is important because it provides information on: the relative sensitivities of different groups of organisms (e.g., invertebrates versus plants); the relative risk associated with different soil-contact exposure pathways (e.g., soil pore water versus ingestion of bulk soil versus interstitial pore vapour); the relative sensitivities among species of the same type (e.g., sensitivity differences among different plant species); and, the relative risk associated with different trophic levels (e.g., microbial function is not impaired but reproduction of arthropods that feed on microorganisms is). The use of a comprehensive *test battery* (i.e., multiple test types and species) decreases the uncertainty associated with the ecological relevance and the accuracy of the toxicity line of evidence. Though not single-species tests themselves, the use of microbial functional tests should also be included in a *battery of toxicity tests* as they provide an indication of the function (e.g., nutrient cycling capability) and microbial diversity in the soil, which provides the basis of the terrestrial food-web.

A further advantage of having more than one species as part of a test battery is that it allows for the possible inclusion of non-standardized, site-specific indigenous species along with standard test species. The toxicity data generated using site-specific indigenous species in conjunction with standard test species reduces the uncertainty and improves the accuracy of the estimation of ecological risk; however, it might lead to highly variable and inconsistent results since the indigenous test species' performance is not known *a priori*. The benefit of using standard methods and associated species is that performance and variability of the test species and method is known.

Table 1. Benefits and limitations of using biological testing in contaminated land management[a]

Benefits of biological testing as a component of contaminated land assessments
• provides a direct estimate of toxicity resulting from short- and long-term soil exposures
• can be a reliable surrogate measure of contaminant bioavailability
• ecologically relevant test species can be selected for use and measurement *endpoints* (such as survival, growth, and reproduction) are meaningful
• results are amenable to *quality assurance/quality control* laboratory measures
• might provide an early indication of ecosystem level change before changes are observed in the field
• response of test organisms is the combined influence of soil physical and chemical characteristics and soil contaminants
• influence of soil physical and chemical characteristics on biological organisms can be differentiated from that of contamination (assuming a well-matched reference soil is also tested)
• provides an estimate of the integrated toxicity of soil contaminated with a complex (and/or unknown) mixture of chemicals or *substances*
• provides an estimate of the toxicity of contaminants for which reliable chemical extraction, analytical methods, or chemical standards do not exist
• provides an estimate of toxicity for contaminants or mixtures of contaminants for which standards or guidelines do not exist
• can generate substance-specific data for the development of soil quality guidelines or standards
• can be an alternative to multiple chemical analyses when the site is, or is suspected to be, contaminated with a complex mixture of compounds
• sites can be mapped with respect to soil contaminant bioavailability in order to refine the accuracy of risk assessments and/or to focus remediation efforts
• multiple sites can be prioritized for remediation according to contaminant bioavailability
• risk through specific soil-contact exposure pathways can be estimated (e.g., soil pore water, soil vapour, or bulk soil)
• risk to soil ecological community function and structure (biodiversity) can be estimated
• risk to floral and faunal communities can be estimated through the use of soil microcosm tests
• use of microcosms can measure the effects of interactions between or among two or more species on soil toxicity (e.g., via added stress on prey species from predation), or can measure the effects of soil toxicity on species interactions (e.g., increased food resources due to reduced competition amongst species)
• bioaccumulation of contaminants in test organisms can be directly measured thereby providing more accurate measurements of risk to organisms at higher trophic levels (e.g., worm-eating birds, seed-eating rodents) through food-web (trophic) transfer

- can be used to evaluate the efficacy of remediation technologies and confirm successful site remediation
- can be used for long-term monitoring of remediation efforts
- use of standardized test methods and procedures can provide comparative data for other, similar, contaminated sites
- if a battery of tests and test species is used, a site-specific *species sensitivity distribution* (SSD) can be generated to derive more realistic and potentially less conservative site-specific benchmarks or remediation objectives
- *chronic* or long-term single-species tests are usually more sensitive compared to field studies; the risk of a false negative is relatively low

Limitations of biological testing in contaminated land assessment

- individual chemical(s) that contribute to the toxicity of a mixture cannot be identified without chemical analysis [although the toxicity identification evaluation approach can assist in identifying the causative substance (or combination of substances)]
- chronic/*sublethal* testing is relatively slow compared with chemical analyses, although rapid toxicity *screening* tests are available
- sensitivity of *acute lethal* tests can be low and can underestimate the long-term effects of contamination; even chronic test data can underestimate toxicity as chronic test durations represent only a fraction of the exposure time at a site
- biological tests can be more expensive compared with routine soil chemical analyses; however, biological tests can be cost-effective in comparison with some inorganic and organic chemical analyses that are expensive or when a complex mixture is present
- test methods have been standardized for a minimal number of species
- soil physical and chemical characteristics (e.g., extreme *pH*, high *salinity* or poor soil *fertility*) can adversely affect test species performance independent of, or in combination with, contaminants
- well-matched reference soils, used to discern the effects of contaminated soil physical and chemical characteristics on test species performance, can be difficult to locate
- biological tests are usually conducted under optimal conditions for the test species; therefore, the species are not exposed to abiotic stressors representative of a natural setting (e.g., drought, extreme temperatures)
- single-species toxicity tests do not typically measure the influence of species interactions (e.g., increased stress due to predation)
- sensitivities of standard test species relative to indigenous species at a site is often unknown (e.g., species at a site might be less sensitive due to adaptation to the presence of the contaminant on site)
- some standard test species or tests can have limited ecological relevance to a site
- there is uncertainty associated with extrapolation of laboratory test results to field effects; therefore, biological testing and soil chemical analyses should ideally be accompanied by ecological field surveys and floral and faunal sampling to increase the confidence of any conclusions made about contaminant toxicity and/or bioaccumulation at a site

[a]Adapted from Thomas *et al.*, 1981; Athey *et al.*, 1987; Greene *et al.*, 1989; Warren-Hicks *et al.*, 1989; NRC, 1995; EC 1999, 2003a; Weeks *et al.*, 2004; Jensen and Mesman, 2006.

In addition, if a test battery comprises a sufficient number of test types (e.g., invertebrate survival and reproduction, plant seedling emergence and growth, and microbial function) and species (e.g., three different species of plants, two earthworm species, different soil arthropod species, and a suite of microbial tests), then a species sensitivity distribution (SSD) can be generated (Posthuma *et al.*, 2002). An SSD can be constructed from site-specific data where detailed knowledge of the relative sensitivities of different soil species across a range of contaminant types and environmental conditions is lacking (EC, 2003a). This distribution can be statistically modelled and then used to provide a threshold effect concentration for site-specific contaminants that can aid in estimating the level of biological impairment relative to soil contaminant concentrations (EC, 2003a). Species sensitivity distributions are widely used in the risk-management of terrestrial and aquatic ecosystems in Canada (CCME, 1999; Gaudet *et al.*, 2001). The specific type and quantity of data needed for an SSD to derive a site-specific guideline or remediation objective are dependent upon the regulatory requirements of the jurisdiction under which the management of the contaminated land falls.

1.2.4 How to Use this Guidance Document

This guidance document has been divided into six sections: Section 1 is the introduction; Section 2 discusses the application of biological testing in contaminated soil assessment; Section 3 provides universal methods and procedures for sampling, handling, transporting, storing, and preparing soils for biological testing; Section 4 provides recommendations and procedures for sampling, handling, transporting, storing, and preparing soils for biological testing that are contaminated with volatile or unstable compounds; Section 5 includes recommendations and procedures for manipulating soil samples to provide soil conditions appropriate for biological testing when these conditions do not occur within samples as collected; and, Section 6 provides specific guidance for sampling, handling, transporting, storing, and preparing soil from *ecozones* in Canada for which some of the universal procedures provided in Section 3 are not applicable or appropriate (e.g., boreal forest, taiga, and tundra ecozones).

Readers of this guidance document unfamiliar with the application of biological testing as part of contaminated site assessments should refer first to Section 2 for recommendations as to which types of biological tests to use, details on those recommended tests, and guidance on how to use the toxicity data generated by the tests to support contaminated land management. Those readers more familiar with the application of biological testing in contaminated land assessment and who are looking for detailed guidance on designing study plans and sampling programs, how to collect and handle soil samples for biological testing, and how to transport, store, and prepare soil prior to biological testing are directed to the universal procedures described in Section 3. Readers who are looking for guidance on how to sample, handle, transport, store, and prepare soils for biological testing that are contaminated with volatile or unstable compounds should refer to Section 4 after first consulting Section 3. Those readers with study objectives that require the manipulation of soil samples for testing beyond the methods described in the universal procedures in Section 3, or, those who require guidance on how to deal with soil that is challenging to prepare should refer to Section 5 after first consulting Section 3. Readers of this document who require guidance on specific procedures necessary to sample, handle, transport, store, and prepare soil from Canada's ecozones for which some of the universal procedures in Section 3 do not apply should refer to the Section 6 after first consulting Section 3.

Application of Biological Testing in Contaminated Soil Assessment

This section recommends biological tests for use in assessing the toxicity of contaminated land. It includes a brief description of each test and the most appropriate application of the test data for contaminated land assessment and management. Tests included are those considered to be useful as screening tools as well as those best suited for more comprehensive *definitive* toxicity evaluations. Case studies are provided (Appendix A) as examples of the use of biological testing in contaminated land management.

2.1 Soil Tests Recommended for Toxicity Assessment of Contaminated Land

The recommended tests are restricted to those for which a standardized national (e.g., Environment Canada) or international (e.g., ISO) method, standard, or test guideline exists. The rationale for recommending standardized test methods is that they are: 1) developed by internationally recognized experts in the field; 2) developed based on the best scientific knowledge and practice at the time; 3) peer-reviewed by a multi-disciplinary scientific community; 4) validated through inter- and intra-laboratory studies involving numerous laboratories to identify the degree of within- and between-lab variability, respectively; and, 5) incorporate rigorous application of quality assurance and quality control (QA/QC) practices (EC, 1999; Spurgeon *et al.*, 2002). In addition to the use of standard methods, it is recommended that biological testing be carried out by laboratories with demonstrated experience and QA/QC procedures for the requested tests. Within Canada, this is most easily achieved by the use of laboratories that are accredited by a third party standardizing organization such as the Canadian Association for Laboratory Accreditation (CALA, formerly CAEAL) or the Standards Council of Canada (SCC). By using data from a consistent set of standardized tests conducted at laboratories by experienced, qualified personnel, the variability associated with procedures and practices can be determined and minimized (EC, 1999; Römbke *et al.*, 2006a).

This is not to suggest that non-standardized methods, designed and conducted following best scientific principles, are not appropriate for use in a contaminated land assessment. Standardized test methods are not available for some test species or assessment endpoints [e.g., trophic level interaction, community level physiological profile (CLPP)]. However, the value of using standard methods is that when conducted by experienced personnel, they generate relevant, reproducible, reliable, and robust data on the toxicity of soils at a contaminated site.

2.1.1 Single-species Toxicity Tests and Microbial Tests

Recommended tests for use for contaminated land assessments include screening, chronic, and definitive single-species toxicity tests and soil microbial tests. Single-species tests are toxicity tests that have two components: the contaminated soil and the test species exposed to that soil. A single test unit may contain one or more individuals of the same species. Common whole-organism test responses (endpoints) include survival, plant seedling emergence, growth, and reproduction. Soil microbial tests evaluate the effect of soil contaminants on the indigenous soil microbial community and its function. Common measurement endpoints in microbial tests are either structural (the composition of the microbial community in the soil), or functional (biologically mediated processes such as nutrient cycling, organic matter breakdown, or soil respiration); these are estimated by measuring microbial activity, biomass, or community structure and diversity (Römbke *et al.*, 2003). Because single-species toxicity tests provide information on the direct toxicity to higher soil-dwelling organisms, and soil microbial tests provide information on the inherent ability of the soil to support these higher levels of organisms and their populations, it is recommended that a combination of single-species

and microbial tests be included in any test battery used in a contaminated land assessment (EC, 2003a). It is currently not recommended that microbial functional and structural tests be used as stand-alone tests because of the difficulty in interpreting the data, unless the data are collected over a long period of time at a frequency that demonstrates a trend or pattern (EC, 2003a).

In addition, it is recommended that Environment Canada test methods, where they exist, be used in preference to similar test methods as these methods were specifically developed using species, test conditions and soil types that are applicable to Canadian ecosystems. They were specifically designed for the assessment of field-collected contaminated soil, rather than for testing soils amended with chemicals, although they can be used for both purposes (EC, 2004a, 2005a, 2007a).

Table 2 provides a brief description of single-species and microbial tests, and applicable references for each test method. The list of single-species and microbial tests is not exhaustive; readers are encouraged to search the literature if the tests listed in Table 2 do not meet site-specific needs.

2.1.2 Soil Microcosm Tests

By using data from multiple single-species tests and microbial structural and functional tests, it is possible to better understand the impact of soil contaminants on ecological communities in the field. This is especially true with improvements in the ecological relevance of newer standardized test methods (EC, 2003b). However, there are still limitations and considerable uncertainty when extrapolating the results from laboratory tests to effects in the field. The advantages of conducting tests in the laboratory are that the exposure and environmental conditions for test species are well-defined and controlled. The limitations are that these tests do not incorporate all abiotic field conditions, and more importantly, they do not account for the effects at the community level by the interactions between different species and populations. Ideally, field experiments, with realistic conditions and spatial scales, would provide a more accurate measurement of adverse effects to ecological receptors at a site. The results from field-scale experiments, however, are highly variable and are confounded by uncontrollable factors such as climatic variations and habitat quality (Bombardier, 2004). A compromise between the control and precision of laboratory studies and the realism of field studies can be met through the use of microcosm tests. Microcosm tests are simplifications of the natural environment and are considered to be any laboratory test that provides data at a biological level of organization greater than single-species (Bombardier, 2004). These tests can be further defined as "controlled reproducible laboratory systems that attempt to simulate the processes and interactions of components in a portion of a natural [terrestrial] ecosystem. The environmental conditions and boundaries are subject to investigator control, and they should include more than one species, preferably at a trophic level higher than microorganisms" (adapted from Morgan and Knacker, 1994 as cited in Bombardier, 2004).

There are a number of advantages to using microcosms in an ecotoxicity assessment. These include:

1. enhanced control over environmental variables;
2. the ability to provide study-specific exposure conditions;
3. lower variability than field studies;
4. they are less costly and labour-intensive than field studies;
5. they provide information on ecosystem function;
6. the influence of species interactions on the effect (including indirect or secondary effects) of the contaminant can be estimated; and,
7. they can be reproduced over time to provide information on ecosystem recovery and/or remediation efficacy (Spurgeon *et al.*, 2002; EC, 2003a; Bombardier, 2004).

The limitations of using microcosms in an ecotoxicity assessment include:

1. they are too simplistic to realistically estimate ecosystem effects;

2. they are too complex to generate direct estimates of toxicity to test species;

3. they tend to be less sensitive than single-species toxicity tests due to ecosystem resiliency and redundancy;

4. they are more costly and labour intensive than single-species tests; and,

5. larger faunal species are excluded (Kuperman *et al.*, 2002; EC, 2003b).

Usually microcosm test systems comprise either an intact soil core or a disturbed bulk soil sample encased in a cylinder of variable dimensions. The preparation and handling procedures of bulk soil samples placed in a column can vary. Contaminants are then typically added to the soil either by topical deposition (e.g., pesticide spray) or by incorporation into the soil. Usually an attempt is made to simulate conditions found in the field. Species are either those naturally occurring in the soil or are introduced test species (e.g., gnotobiotic species). Indigenous or gnotobiotic test species can include microorganisms, mesoinvertebrates (e.g., collembola, mites), macroinvertebrates (e.g., earthworms, beetles), fungi, and plants. Ecosystem-level endpoints are then measured. These can include:

1. microbial respiration;
2. nutrient cycling;
3. microbial population changes;
4. changes in microbial biomass;
5. changes in decomposition rates of litter and/or soil organic matter;
6. changes in the population of micro- and mesoinvertebrates;
7. survival, growth, and reproduction of micro- and macroinvertebrates;
8. plant emergence, growth, and yield.

Recovery rates of ecosystem function and structure after stress and contaminant degradation or loss through leachate can also be measured (Linder *et al.*, 1992; EC, 2003a; Bombardier, 2004).

As stated, microcosm tests can be conducted with *unconsolidated soil samples* or with consolidated (intact) soil cores, depending on the study objective. The use of intact soil cores provides more ecological realism than using unconsolidated soil samples; however, data are more variable and require more field collection effort to overcome this variability (more field replicates). Relatively standardized microcosm methods using intact soil cores (e.g., terrestrial model ecosystems, or TMEs) have been developed (e.g., ASTM, 1993) and successfully used in the evaluation of the fate and effect of crop protection products on soil organisms (Linder *et al.*, 1992). Guidance for sampling soil cores for TMEs exists and good TME-to-field correlations have been demonstrated for the effect of pesticides on biological community structural endpoints (Bombardier, 2004; Knacker *et al.*, 2004; Nikolakis *et al.*, 2007; Schaeffer *et al.*, 2008). Terrestrial model ecosystems have been validated in an interlaboratory-comparison study at four European sites by comparing the influence of different soil types and organism groups in the TMEs and by comparing field results at the sites from where the TMEs were extracted (Weyers *et al.*, 2004; Moser *et al.*, 2007). However, despite the fact that TMEs have been used for the assessment of contaminated soils, particularly in the Netherlands (Kools *et al.*, 2009), currently they are semi-field methods which are primarily used in the assessment of pesticides applied in the environment. As such, TMEs lend themselves better to experimental designs in which the contaminant of concern is added to the test system after it has been established.

Although a wide variety of microcosm test systems exist, there is no standard test design for microcosms. However, based on the results of a recent international Society of Environmental Toxicology and Chemistry (SETAC) *Environmental Risk Assessment* of Pesticides in Soil (PERAS) workshop (Coimbra, Portugal, October 2007), an initiative was started to assess the potential use of TMEs as a standard higher-tier test method (Schaeffer *et al.*, 2008). Nevertheless, there is a critical lack of guidance for the interpretation of results from microcosm tests. The advantages accrued by the use of biological testing in contaminated land assessments are predicated on the fact that the test methods are standardized, test results are reproducible, and the use and interpretation of the data are not ambiguous to both the site assessor and regulatory agency. Microcosms in their current state of

Table 2. Soil tests applicable for toxicity assessment of contaminated land*

Test type	Test description	Duration	Biological endpoint(s)	Test species	Reference(s)
Single-species toxicity tests: Screening tests					
Acute earthworm survival	Lethal: measures the *acute toxicity* of soil to adult earthworms	14 d	Survival	*Eisenia andrei, Eisenia fetida* *Lumbricus terrestris*	EC, 2004a** ISO, 1993a (*Eisenia* sp. only) methods also published in scientific literature (e.g., Greene *et al.*, 1989)
Earthworm avoidance	Sublethal: measures the avoidance of adult earthworms to contaminated soil	48 to 72 h[a]	Avoidance behaviour	*Eisenia andrei. Eisenia fetida* *Lumbricus terrestris*	EC, 2004a** ISO, 2008a (*Eisenia* sp. only)
Phytotoxkit™	Lethal/sublethal: measures seedling germination, emergence, shoot and root length	3 d	Germination, shoot and root length	*Sorghum saccharatum, Sinapis alba, Lepidium sativum*	No standard method but commercialized kit available[b]
Seedling emergence	Lethal: measures seedling emergence	5 d	Emergence	*Lactuca sativa* L.	ISO, 2005d
Seedling root elongation test	Lethal/sublethal: measures root length	5 to 7 d[a]	Root length	Many species. Refer to ISO method for details	ISO, 1993b; methods also published in scientific literature [e.g., Greene *et al.*, 1989 (*L. sativa* only)]
Juvenile insect (herbivorous beetle) survival	Lethal: measures the acute toxicity of soil or food to insect larvae	10 d	Survival	*Oxythyrea funesta*	ISO, 2005c
Acute nematode[c] survival	Lethal: measures the acute toxicity of soil to subadult nematodes	24 h	Survival	*Caenorhabditis elegans*	ASTM, 2001

Test type	Test description	Duration	Biological endpoint(s)	Test species	Reference(s)
Single-species toxicity tests: Chronic/definitive tests					
Earthworm reproduction	Sublethal: measures the chronic survival and reproductive capability of adult earthworms	56 to 63 d	Adult 28 to 35-d survival. Number of progeny produced Progeny wet and dry mass	*Eisenia andrei*	EC, 2004a**
Earthworm reproduction	Sublethal: measures the chronic survival and reproductive capability of adult earthworms	56 d	Adult 28-d survival Change in adult biomass. Number of cocoons produced	*Eisenia andrei* *Eisenia fetida*	ISO, 1998
Plant emergence and seedling growth	Sublethal: measures seedling emergence and early growth	14 to 21 d[a]	Emergence. Shoot and root length. Shoot and root wet and dry mass Phytotoxicity	Many species. Refer to method for details	EC, 2005a**
Plant emergence and seedling growth	Sublethal: measures seedling emergence and early growth	14 to 21 d[a] minimum	Emergence. Seedling wet and dry mass	Many species. Refer to method for details	ISO, 1995
Plant life-cycle test	Sublethal: measures seedling growth and yield	35 to 64 d[a]	Biomass Shoot length Number of seed pods Number of flowers	*Brassica rapa* L. *Avena sativa*	ISO, 2005f
Collembola[d] reproduction	Sublethal: measures adult chronic survival and reproduction	21 to 28 d[a]	Adult survival Number of progeny produced	*Folsomia candida, Orthonychiurus folsomi, Folsomia fimetaria*	EC, 2007a**
Collembola reproduction	Sublethal: measures adult chronic survival and reproduction	28 d	Adult survival Number of progeny produced	*Folsomia candida*	ISO, 1999

Test type	Test description	Duration	Biological endpoint(s)	Test species	Reference(s)
Enchytraeid[e] reproduction	Sublethal: measures the chronic survival and reproductive capability of adult potworms	28 to 42 d[a]	Adult 21-d survival Number of progeny produced	*Enchytraeus albidus* *Enchytraeus crypticus*	ISO, 2004a ASTM, 2004
Oribatid mite[f] reproduction	Sublethal: measures the chronic survival and reproduction capacity of adult soil mites	28 d	Adult 28-d survival Number of progeny produced	*Oppia nitens*	No standard method; methods published in scientific literature (e.g., Wiles and Krogh, 1998; EC and SRC, 2007; Princz, *et al.*, 2010)
Predatory mite[g] reproduction	Sublethal: measures the chronic survival and reproductive capicity of adult soil mites	16 d	Adult survival Number of progeny produced	*Hypoaspis aculeifer*	OECD, 2008
Snail growth and survival test	Sublethal: measures the chronic survival and growth of adult snails	28 d	Adult survival Adult wet tissue mass	*Helix aspersa*	ISO, 2006b
Microbial toxicity tests (summaries modified from Rahn, 2008) **Soil microbial biomass tests**					
Fumigation/ extraction test	Chloroform used to extract microbial carbon from soil; carbon (C) extracted then measured analytically; does not differentiate between live and dead microbial C	Up to 72 h	Total microbial carbon (biomass)	Indigenous soil community	ISO, 1997a
Soil microbial activity tests					
Soil microbial respiration/Basal respiration (BR)	Measures the basal CO_2 production of a soil sample as an indication of the activity of live/active microorganisms	3 to 5 d	Total CO_2 production	Indigenous soil community	ISO, 2002d
Substrate induced respiration (SIR)	Measures CO_2 production after the addition of glucose (or other substrate) to soil; measures activity of live/active microorganisms	$\geq$ 28 d	Total CO_2 production	Indigenous soil community	ISO, 1997b ISO, 2002c

Test type	Test description	Duration	Biological endpoint(s)	Test species	Reference(s)
Soil nitrification/ denitrification	Measures the nitrification/denitrification by microorganisms in soil as an indication of soil fertility and nutrient cycling capabilities	≥ 28 d	Measures nitrogen (N) species (NO_3^-) extracted from soil	Indigenous nitrifying/ denitrifying soil populations	ISO, 2004b ISO, 1997c
Bait lamina test	Measures the feeding activity of soil invertebrates and/or microorganisms; thin plastic sticks with holes filled with a standard carbon substrate are placed into soil; can be used as a laboratory or field test	4 to 21 d depending on climatic conditions	Number of empty holes per stick (number of empty holes corresponds to amount of substrate eaten)	Indigenous soil community and/or exogenous soil invertebrates added to test unit	No standard method; methods published in scientific literature for lab application (e.g., Beyaert and Fox, 2008)
Arthrobacter test	Bacterial soil-contact toxicity test: measures dehydrogenase activity following 2-h exposure to substrate	6 h	Fluorimetric measurement of resorufin activity	*Arthrobacter globiformus*	ISO, 2008b
Soil microbial diversity and community structure					
Community level physiological profile (CLPP)	Soil diluted with water and inoculated on microplates containing different types of carbon sources; differential utilization of carbon provides indication of microbial community diversity; unculturable organisms not detected	2 h to 5 d	Average well colour development in microplates (well colour development indicates substrate use)	Indigenous soil community	No standard method but commercialized CLPP plates available[h]
Denaturing gradient gel electrophoresis (DGGE)	Extraction of DNA from soil samples. DNA is amplified by specialized PCR[i] and analyzed by polyacrylamide gel electrophoresis. Characteristics of DNA denatured on gel corresponds to microbial community structure and diversity	2 h to 1 to 2 d	Number, distribution and intensity of PCR[i]-amplified DNA bands on gel	Indigenous soil community	No standard method; methods published in scientific literature (e.g., Topp *et al.*, 2008)

Test type	Test description	Duration	Biological endpoint(s)	Test species	Reference(s)
Soil enzyme assays	Analysis of enzyme activity in soil; microbial and extracellular enzyme activity drives many soil functions and influences soil fertility; assays provide an indication of community function; enzymes commonly targeted are: phosphatases, sulphatases, glycosidases, amidohydrolases and arylamidases	Variable: depends on the endpoint measured	Colorimetric and fluorescence endpoints used to measure enzyme activity	Indigenous soil community	ISO, 2005g ISO, 2005h

*See Table 4 in Linder *et al.* (1992) for an evaluation of the applicability of specific tests to different habitat types

**Preferred test. Environment Canada test methods, where they exist, are recommended over other, similar standard methods for assessment of contaminated land in Canada

[a]Test duration differs depending on the test species

[b]Phytotoxkits are available from MicroBioTests Inc., Kleimoer 15, 9030 Mariakerke (Gent), Belgium www.microbiotests.be

[c]Nematodes used in soil toxicity methods are free-living (non-parasitic) soil roundworms

[d]Collembola, also known as springtails, are small soil-dwelling arthropods

[e]Enchytraeids, also known as potworms, are small white segmented worms that feed on organic material in soil

[f]Oribatid mites are soil and humus mites that feed on fungi, algae, organic debris, and dead collembola. Some are predaceous and feed on nematodes and *microfauna*.

[g]Predatory mites are mites that feed on other *mesofauna* (e.g., collembola) or insect larvae

[h]CLPP plates are available from commercial suppliers (e.g., Biolog Inc., 21124 Cabot Blvd., Hayward, California 94545, USA www.biolog.com.)

[i]PCR – polymerase chain reaction

development do not yet meet these requirements. Further detail on soil microcosms and discussions of the advantages and limitations of the various methods can be found in Linder (1992), Kuperman *et al.* (2002), Spurgeon *et al.*, (2002), EC (2003a), Römbke *et al.* (2003); a comprehensive review is found in Bombardier (2004) on both microcosms in general and TMEs in particular.

2.2 *Use of Biological Data in Contaminated Land Management*

As mentioned in earlier sections of this document, in the past biological (toxicity) testing has normally been integrated into contaminated land assessment at higher tiers of a risk assessment (e.g., Tier 2 or 3). However, biological testing has been used more recently at a Tier 1 (screening) level of the risk assessment process. Regardless of what level of biological testing is used in a site assessment, it is critical that the tests conducted are relevant and the test data are appropriately interpreted and applied. This means that for every biological test conducted:

- the questions to be answered must be clearly articulated;
- the data needed to answer these questions must be known;
- there must be a thorough understanding of how the data generated by the tests can answer these questions (including an understanding of the limitations of the data generated);
- the stage in the contaminated land assessment or management that the data will be applied and used should be known;
- the potential for iterative testing should be anticipated (e.g., positive acute screening test results will be followed by more focused definitive tests); and,
- the acceptable level of test quality must be established before tests are conducted.

These issues need to be discussed at the earliest planning stages of an assessment and discussions should include not only the site assessor but the site manager, regulators, ecologists, statisticians, ecotoxicologists, and the ecotoxicity testing laboratory(s) involved in the contaminated site project.

Table 3 provides examples of the use of site-specific toxicity data within different stages of a contaminated land management. The uses of the data described in Table 3, however, are not exhaustive; they assume a tiered assessment approach and are subject to modification depending on the specific objectives of the assessment and the conditions at the site. Guidance on contaminated land assessments is found in CCME (1996b). Case studies describing the use of biological testing in contaminated site management are provided in Appendix A.

The value of using the same type of test (e.g., invertebrate reproduction) with more than one type of species (e.g., earthworm and collembola) in a contaminated site assessment test battery is that exposure to contaminants through different pathways (e.g., soil pore water, soil air) is evaluated, and that the sensitivity of soil dwelling organisms from different trophic *guilds* and taxonomic groups is measured. The physical and chemical properties of contaminants and environmental media strongly influence the bioavailability of contaminants to soil organisms. Assessing organisms that are exposed through different soil contact pathways, therefore, provides a more complete assessment of the risk to all soil-dwelling organisms at a site. Including organisms in a test battery from different trophic levels also provides more information on the possible effect of a contaminant(s) through food web transfer at a site. Table 4 provides a summary of the possible exposure pathways and trophic levels represented by common test organisms exposed to contaminants through the soil-contact pathway.

Table 3. Examples of the use of biological test data in contaminated land management

Test type	Question to be answered	Use(s) of data in contaminated land management
Screening tests		
Acute invertebrate survival	Is soil highly toxic (adults cannot survive short periods of exposure)?	To screen a site for highly toxic areas either before or after chemical characterization (Tier 1, 2). Use in Tier 1: can help focus chemical sampling efforts Use in Tier 2: can identify areas that are highly toxic and will likely require remediation Use in Tier 2: can help focus efforts on less toxic areas suitable for Tier 3 sublethal testing and risk assessment
Plant emergence	Is soil highly toxic (seeds cannot germinate/emerge)?	To screen a site for highly toxic areas either before or after chemical characterization (Tier 1, 2). Use in Tier 1: can help focus chemical sampling efforts Use in Tier 2: can identify areas that are highly toxic and will likely require remediation Use in Tier 2: can help focus efforts on less toxic areas suitable for Tier 3 sublethal testing and risk assessment
Phytotoxkit™	Is soil highly toxic (seeds cannot germinate, or can germinate but growth is inhibited following short period of exposure)?	To screen a site for highly toxic areas either before or after chemical characterization (Tier 1, 2). Use in Tier 1: can help focus chemical sampling efforts Use in Tier 2: can identify areas that are highly toxic and will likely require remediation Use in Tier 2: can help focus efforts on less toxic areas suitable for Tier 3 sublethal testing and risk assessment
Short-term plant root elongation	Is soil highly toxic (seeds can emerge, but growth is inhibited following short period of exposure)?	To screen a site for highly toxic areas either before or after chemical characterization (Tier 1, 2). Use in Tier 1: can help focus chemical sampling efforts Use in Tier 2: can identify areas that are highly toxic and will likely require remediation Use in Tier 2: can help focus efforts on less toxic areas suitable for Tier 3 sublethal testing and risk assessment
Chronic/ Definitive tests		
Earthworm avoidance	Is soil too toxic to support invertebrate populations? [avoidance is more sensitive than survival and often shows	To screen a site for potentially sublethally toxic areas either before or after chemical characterization (Tier 1, 2). Use in Tier 1: can help focus chemical sampling efforts Use in Tier 2: can help focus efforts on less toxic areas suitable for Tier 3 sublethal testing and risk

	similar effects to those found in reproduction (ESG, 2000; Garcia *et al.* 2008; Hund-Rinke and Wiechering, 2001; Hund-Rinke *et al.* 2003)]	assessment
Invertebrate reproduction	Is soil too toxic to support invertebrate populations?	To evaluate the potential for contaminated soil to inhibit invertebrate chronic survival and reproduction (Tier 2, 3). Use in Tier 2,3: can identify site soil (concentrations of contaminant in a site soil) that inhibits invertebrate chronic survival and reproduction Use in Tier 2,3: can identify soil characteristics that modify bioavailability of contaminant Use in Tier 2,3: can be used to derive (in part) a site-specific standard and/or remedial objective for the site Use in Tier 2,3: can identify the efficacy of bioremediation technologies and/or site remediation *Risk management*: can be used for long-term monitoring of a remediated site
Emergence and early seedling growth	Is soil too toxic to allow plant establishment and growth?	To evaluate the potential for contaminated soil to inhibit emergence and longer-term growth of plants (Tier 2, 3). Use in Tier 2,3: can identify site soil (concentrations of contaminant in a site soil) that inhibits plant emergence and growth Use in Tier 2,3: can identify soil characteristics that modify bioavailability of contaminant Use in Tier 2,3: can be used to derive (in part) a site-specific standard and/or remedial objective for the site Use in Tier 2,3: can identify the efficacy of bioremediation technologies and/or site remediation Risk management: can be used for long-term monitoring of a remediated site
Plant life-cycle tests	Is soil too toxic to support sustainable plant populations? Does contamination in soil adversely affect crop yield?	To evaluate the potential of contaminated soil to inhibit emergence, long-term growth, and reproduction of plants (Tier 3). Use in Tier 3: can identify site soil (concentrations of contaminant in a site soil) that inhibits plant emergence and growth and reproduction Use in Tier 3: can identify site soil (concentrations of contaminant in a site soil) that inhibits crop yield Use in Tier 3: can identify soil characteristics that modify bioavailability of contaminant Use in Tier 3: can be used to derive (in part) a site-specific standard and/or remedial objective for the site Use in Tier 3: can identify the efficacy of bioremediation technologies and/or site remediation Risk management: can be used for long-term monitoring of a remediated site
Plant or invertebrate	Does the contaminant(s)	To evaluate the potential of contaminants in soil to bioaccumulate in plant or invertebrate tissue

bioaccumulation [other animals (e.g., moles, foxes) can also be used to estimate bioaccumulation in wildlife]	bioaccumulate in plants or invertebrates, thereby posing a risk to organisms that feed on them?	(Tier 3). Use in Tier 3: can identify site soil (concentrations of contaminant in a site soil) that results in the bioaccumulation of significant levels of contaminant in plant and invertebrate tissue (i.e., a site-specific bioaccumulation factor) Use in Tier 3: can identify soil characteristics that modify bioavailability of contaminant Use in Tier 3: can be used to estimate the site-specific relationship between soil contaminant level, toxicity, and tissue/whole organism contaminant concentrations Use in Tier 3: can be used to derive (in part) a site-specific standard and/or remedial objective for the site Use in Tier 3: can be used to refine the dietary exposure estimates of ecological receptors that feed on plants and invertebrates present on the site
Microbial tests		
Soil respiration	Is the contamination adversely affecting general soil function?	To evaluate the potential for contaminated soil to inhibit normal microbial activity in the soil (Tier 2, 3). Use in Tier 2,3: can identify site soil (concentrations of contaminant in a site soil) that inhibits general microbial activity Use in Tier 2,3: can identify soil characteristics that modify the toxicity of contaminants to microbial function Use in Tier 2,3: can be used to derive (in part) a site-specific standard and/or remedial objective for the site Use in Tier 2,3: can identify the efficacy of bioremediation technologies and/or site remediation Risk management: can be used for long-term monitoring of a remediated site
Nutrient mineralization by soil microorganisms	Is the contamination adversely affecting nutrient cycling function of the soil?	To evaluate the potential for contaminated soil to inhibit the normal nutrient cycling activities (e.g., nitrogen and carbon mineralization etc.) of the soil microbial community (Tier 2, 3). Use in Tier 2,3: can identify site soil/concentrations of contaminant in a site soil that inhibit nutrient cycling Use in Tier 2,3: can identify soil characteristics that modify the toxicity of contaminants to microbial function Use in Tier 2,3: can be used to derive (in part) a site-specific standard and/or remedial objective for the site Use in Tier 2,3: can identify the efficacy of bioremediation technologies and/or site remediation Risk management: can be used for long-term monitoring of a remediated site
Bait Lamina	Is the contamination adversely affecting the abundance and activity of soil micro-, meso- and *macrofauna*?	To evaluate the potential for contaminated soil to inhibit the normal feeding activity of soil fauna on soil organic matter (Tier 2, 3). Use in Tier 2,3: can identify site soil (concentrations of contaminant in a site soil) that inhibits feeding activity of soil fauna Use in Tier 2,3: can identify soil characteristics that modify the toxicity of contaminants to soil micro-, meso- and macrofaunal feeding behaviour Use in Tier 2,3: can be used to derive (in part) a site-specific standard and/or remedial objective for the site

		Use in Tier 2,3: can identify the efficacy of bioremediation technologies and/or site remediation Risk management: can be used for long-term monitoring of a remediated site
Soil microbial DGGE	Is the contamination adversely affecting soil diversity and structure, and hence potential community or resiliency, of soil microorganisms?	To identify the effect of soil contamination on the diversity of the microbial community in the soil (Tier 2, 3). Use in Tier 2,3: can identify site soil (concentrations of contaminant in a site soil) that decreases soil microbial diversity Use in Tier 2,3: can identify soil characteristics that modify the effect of contaminants on the diversity of soil microorganisms Use in Tier 2,3: can be used to derive (in part) a site-specific standard and/or remedial objective for the site Use in Tier 2,3: can identify the efficacy of bioremediation technologies and/or site remediation Risk management: can be used for long-term monitoring of a remediated site
Soil microbial CLPP	Is the contamination adversely affecting the microbial biodiversity, and hence potential community resiliency, of soil microorganisms?	To identify the effect of soil contamination on the diversity of the microbial community in the soil (Tier 2, 3). Use in Tier 2,3: can identify site soil (concentrations of contaminant in a site soil) that decreases soil microbial diversity Use in Tier 2,3: can identify soil characteristics that modify the effect of contaminants on the diversity of soil microorganisms Use in Tier 2,3: can be used to derive (in part) a site-specific standard and/or remedial objective for the site Use in Tier 2,3: can identify the efficacy of bioremediation technologies and/or site remediation Risk management: can be used for long-term monitoring of a remediated site
Soil enzyme assays	Is the contamination adversely affecting soil functions?	To identify the effect of soil contamination on soil function (Tier 2, 3). Use in Tier 2,3: can identify site soil (concentrations of contaminant in a site soil) that decreases soil function Use in Tier 2,3: can identify soil characteristics that modify the effect of contaminants on the function of soil Use in Tier 2,3: can be used to derive (in part) a site-specific standard and/or remedial objective for the site Use in Tier 2,3: can identify the efficacy of bioremediation technologies and/or site remediation Risk management: can be used for long-term monitoring of a remediated site

Table 4. Generic contaminant exposure pathways for different types of commonly used biological test organisms

Organism	Trophic level and food source	Exposure pathway
Micro-organism	Producers Usually derive energy from sunlight (e.g., algae)	Diffusion or uptake from bulk soil, pore water, soil, and air
	Consumers Feed on inorganic and *organic soil* material	Diffusion or uptake from bulk soil, pore water, soil air, food
Plants	Producers Energy derived from sunlight	Diffusion from bulk soil Uptake from soil pore water and soil air
"Soft-bodied" invertebrates		
Nematodes	Consumers (detrivores, herbivores, carnivores) Feed on bacteria, protozoa, fungi, plant roots and/or exudates, other nematodes	Ingestion of food, pore water Inhalation of soil air
Enchytraeids	Consumers (detrivores) Feed on organic matter, litter, fungi, and microorganisms	Ingestion of food, pore water Inhalation of soil air Dermal contact of bulk soil, pore water
Snails	Consumers (detrivores and herbivores) Feed on fungi, organic matter, living plants	Ingestion of food, pore water
Earthworms	Consumers (detrivores) Feed on litter, organic matter, fungi, and microorganisms	Ingestion of bulk soil, food, pore water Inhalation of soil air Dermal contact of bulk soil, pore water
"Hard-bodied" invertebrates		
Collembola	Consumers (detrivores) Feed on fungi, organic matter, nematodes, bacteria	Ingestion of pore water, food Inhalation of soil air
Mites	Consumers (detrivores, carnivores) Feed on fungi and organic matter, collembola, nematodes, and other meso- and micro-fauna	Ingestion of pore water, food Inhalation of soil air

Section 3

Universal Procedures for Sampling, Handling, and Preparing Soils for Biological Testing

3.1 Overview

This section provides guidance on the sampling, transport, storage, handling, and preparation of soils for biological testing. It describes the steps required to prepare a study plan and includes information that is useful for conducting a site survey, how to select a *sampling strategy*, recommended soil collection procedures, a description of sampling devices, sample transport considerations, storage recommendations, and soil preparation techniques. Readers who require guidance on the specific procedures necessary to sample, handle, transport, store, and prepare soil contaminated with volatile or unstable compounds should also refer to Section 4. Those readers who have specific study objectives that require the manipulation of soil samples beyond soil preparation methods described in Section 3, or, those who require guidance on how to deal with soil that is challenging to prepare should also refer to Section 5. Readers who require guidance on the specific procedures necessary to sample, handle, transport, store, and prepare soil from Canada's ecoregions, in particular boreal, taiga, and tundra ecozones, should also refer to Section 6.

It is assumed that users of this guidance are collecting and handling soils for biological testing as part of a higher level site or risk assessment and therefore the chemical contamination and soil physical and chemical properties of the land under investigation have already been well-characterized. Unless otherwise stated, it is assumed that the general nature, concentrations, distribution, and degree of variability of the contaminants and soil physical and chemical characteristics at the site are reasonably well-delineated. This guidance can apply to circumstances in which chemical analyses are limited due to the complexity of the contamination at the site or the unusual nature of the contaminants (e.g., mixtures of polymers, spills of specialty chemicals, etc.); however, the guidance provided in this section is not intended to be applicable to soil sampling for chemical analyses, although there are some similar recommendations and procedures.

3.2 Study Objectives

Defining the objectives and goals is the first and most important step in developing a study plan that will incorporate biological testing into contaminated land assessment and management. To determine the study objectives, two critical questions must be posed:

- What questions need to be answered?
- What data are needed to answer the questions?

It is advisable that the study objectives be developed and agreed to by all stakeholders, including regulators, site assessors, and the proponent during early stages of the site assessment.

By answering these two fundamental questions at the outset of a study, the investigator can better:

- select the proper tests to meet the study objectives;
- select the proper test species to meet the study objectives;
- determine the sampling strategy that will optimize the utility and interpretation of the data and subsequent understanding of the site;
- plan the appropriate methods (including QA/QC procedures) for sample collection, handling, transport, storage, and preparation;
- understand what background data already exist about the site that do not need to be collected again;

- identify the supplementary information that also needs to be collected to help interpret the test results (identify the data gaps); and,
- determine the most cost-effective approach to maximize the quality and usefulness of the data obtained (Greene *et al.*, 1989; ISO, 2002c).

The identification of study objectives is assisted by defining the assessment and measurement endpoints for the project. Assessment endpoints are important environmental qualities or values to protect, for example, an assessment endpoint could be soil fertility (EC, 2003a). Measurement endpoints are used to evaluate the impacts of contaminated soil on assessment endpoints (EC, 2003a). Many different measurement endpoints can be evaluated for one assessment endpoint. For example, measurement endpoints for soil fertility could include nitrogen and carbon cycling, invertebrate feeding activity and plant growth and yield (EC, 2003a; Jensen *et al.*, 2006b; Mesman *et al.*, 2006). Individual assessment and measurement endpoints can be identified as study objectives.

Examples of other study objectives include:

- generate a species-sensitivity distribution for the site;
- identify the ecological receptors most at risk at the site;
- derive threshold effect concentrations or remedial objectives for the site;
- generate other toxicity or bioavailability data needed by risk assessors;
- characterize the toxicity of the site in order to design site remediation measures;
- evaluate the efficacy of site remediation technologies; and,
- provide long-term monitoring data for the site.

3.3 Study Plan

Once the study objectives have been defined, the next step is to develop the study plan. A study plan provides specific guidance for the methods and strategies for sample collection and the procedures required to ensure that all *data quality objectives* are met (EC, 1994). A study plan must be carefully thought out and all personnel who will contribute to the project should be involved as early in the process as possible, preferably at the study plan design stage. These personnel typically include the project/site manager, a statistician, a soil scientist, a field biologist, a risk assessor, ecotoxicologists(s) who will be conducting the biological testing, and any other personnel with the requisite expertise (Athey *et al.*, 1987). In addition, it is advisable that the selection of biological tests, endpoints and test species be discussed with, and agreed to, by all stakeholders, including regulators, site assessors, and the proponent during the study plan design stage.

The following information should be included in any study plan:

- data quality objectives
- definition of the *study area*
- background data collection
- site survey
- selection of sampling locations
  - sampling strategy
  - identification of sampling locations
- *sample size* (volume, mass)
- number of samples
- quality assurance and quality control protocols
- sampling plan
  - sample plan checklist
  - field measurements and observations
  - field notes checklist
  - sampling procedures
  - sampling devices
  - sample preparation and/or storage in the field
  - sample transport and documentation
  - contingency plans
  - sample QA/QC procedures
  - environmental, health, and safety procedures
- sample receipt (at the laboratory)
- sample storage
- sample preparation
- sample manipulation
- biological test design
- use of biological test data
- sample archiving and disposal

Table 5 provides a checklist of the questions that need to be answered when planning for biological testing of contaminated sites. Detailed guidance or descriptions of the items listed in Table 5 are provided in Sections 3.2 to 3.11.

3.3.1 Identification of Data Quality Objectives (DQOs)

Data quality objectives are the full set of constraints needed to design a study. They are statements that provide definitions of the confidence required to make conclusions from the study data and determine the degree of total variability that can be tolerated in the data (CCME, 1993a; 1996b). Variability can be a reflection of uncertainty or error, or it can be inherent to a system and known. Another way to describe DQOs is that they indicate the overall level of uncertainty that a decision-maker is willing to accept in results derived from environmental data (USEPA, 1986). Data quality objectives can be qualitative (e.g., rules to follow if a sample cannot be taken at the exact planned location) or quantitative [e.g., the standard deviation desired for the mean toxicity of a sample(s)]. Setting the DQOs will impact the size and cost of the study as they will dictate the site location, number of samples, size of samples and extent of biological testing (CCME, 1993a). Sometimes a compromise will have to be reached between the desired DQOs and the cost of the sampling and testing.

CCME (1993a) provides a generic description of the steps involved in developing DQOs, including the following summary:

- state the problem to be resolved
- identify the decision to be made
- identify the inputs to the decision
- narrow the boundaries of the study
- develop a decision rule
- develop uncertainty constraints
- optimize design for obtaining data

Considerations that should be part of the above DQO process include:

- timeline
- budget
- access to site
- climate
- weather
- activity on site (human and ecological)
- access to background data
- identification of decisions to be made
- identification of inputs into decisions
- acceptable levels of uncertainty for the decision-maker
- acceptable levels of variability in the data for the decision-maker
- decision rules (e.g., contingency plans if a sample cannot be taken at the planned location because of the presence of rocks or a tree)
- regulatory requirements of the data
- anticipated challenges (e.g., procedures to follow when collecting soils contaminated with volatile contaminants, how to sample very shallow soils)

3.3.2 Definition of the Study Area

The study area refers to a locale that contains the *study site* to be assessed, as well as adjacent landscape that might influence the conditions of the study site. The study site refers to a smaller zone within the study area that is to be assessed, monitored, or from which the samples are to be collected. The boundaries of both the study area and study site should be clearly delineated and documented on maps and by other means [e.g., global positioning system (GPS) coordinates].

3.3.3 Background Data Collection

Conducting a desktop search of background data about the site provides much information pertinent to the study objectives, study plan, and sampling plan. If the study is being conducted as part of a higher tier risk or site assessment, there should already be a significant amount of data acquired about the site. These data would have been collected during the initial investigations conducted as part of the screening stages of the study, and as such, would also include site-specific data collected as part of those site investigations. If no background information has yet been collected for the site, the following list provides examples of pertinent background data for the site (CSA, 2001):

- location
- access
- infrastructure on or near the site (e.g., conduits, sewers, underground storage tanks, buried wires, gas and water pipes)

Table 5. Study plan checklist*

- What are the study objectives?

- What are the data quality objectives (DQOs)?
- What will be done if the DQOs are not met (resample or revise DQOs?)?

- What constitutes the study site?
- What is the site location?
- What are the site boundaries?
- What regulatory jurisdiction does the site fall under?

- Have arrangements been made to collect samples from the site?
- What are the contingency plans in case not all sites can be sampled?

- What biological tests are specified?
- What species are to be tested?
- How many species/tests are specified?
- How many types of tests are specified?
- What specific biological test methods and method QA/QC protocols are required?
- What is the maximum number of samples that can be run together as a single toxicity test?
- Does the *sampling design* need to be changed to accommodate the toxicity laboratory maximum test load?

- Which chemical analyses need to be conducted on soil samples collected for biological testing?
- When do these analyses need to be conducted (e.g., following sampling, following *test soil* preparation etc.)?

- Are all the contaminants listed or known? Do they all need to be known?
- Is the chemical characterization at the site adequate to design the sampling study for biological testing [including the selection of the reference soil(s)], for example, are the distribution and variability of contamination in soil known?
- If it is not, will re-sampling be required?

- Can a suitable reference soil(s) be obtained?
- What is the alternative in the event that a suitable reference soil cannot be obtained?

- What equipment is needed?
- Is any kind of specialized equipment, sampling or testing procedures needed (e.g., if soils contain volatile contaminants)?

- Are soil *samplers* experienced in the required sampling procedures?
- If not, how will samplers be trained/experienced personnel be obtained to sample?

- Are there any special health and safety concerns at the site?
- What preparations are in place to avoid environmental contamination due to sampling procedure (e.g., equipment decontamination)?
- Are there any special health and safety concerns for laboratory personnel?
- Will the samples be collected by horizon or by depth?
- What is the maximum depth for sampling?

- Which sampling strategy will be used (e.g., random, systematic, judgmental, etc. or a combination)?
- Will the type of sampling meet the DQOs?

- Which type of data analyses will be used (e.g., geostatistical, hypothesis testing, regression)?
- Will the data analyses meet the DQOs?
- Is the sampling approach compatible with the data analysis methods?

- How many sites are there?
- How many sampling locations are there?
- How many samples are needed (per site and per *sample location*)?
- Will the samples be consolidated or unconsolidated?
- Will the unconsolidated samples be composited?
- How long will it take to collect all the samples at a site and/or location?
- What are the contingency plans if samples cannot be collected within the estimated time?
- What is the order of sampling sites/locations? Will the order be randomized?
- How many tests are specified?
- What volume of soil (what sample size) is needed per sample?
- How many reference samples are needed?
- How much reference soil is needed for all tests and analyses?
- Which types and how many QA/QC samples are needed to meet the DQOs? (e.g., *replicate samples*)
- Will any supplementary samples be taken?

- Which field measurements will be taken? (e.g., *in-situ* toxicity tests, vegetation surveys)

- How will the soil be handled on-site (e.g., sieved, dried, homogenized)?
- How will the soil be transported?

- How will the soil be stored and handled at the laboratory?
- What is the holding time for the soil (e.g., should testing be performed immediately)?
- Are there any restrictions to soil preparation (e.g., no air drying, no *sieving*, etc.)?
- What is the contingency if soil preparation restrictions must be breached in order to test the soil?
- Are any soil manipulations required?
- What is the alternative if soil manipulations are not successful?

- Does the testing laboratory have appropriate and adequate control soil for assessing the test validity during toxicity testing? (If not, collection of a control soil may become part of the sampling program).

- Will the soil samples need to be stored for another phase of testing?
- Should the biological test material (e.g., plant tissue, invertebrate tissue) from the tests be stored for future analyses?
- How will the samples be disposed?

*Adapted from CCME, 1993a

- past and current utilization of the site (e.g., past disposal activities, chemical storage)
- potential health and safety hazards on the site (e.g., buried wires, gas pipes)
- location of contaminant source relative to site
- climate of the study area
- ecoregion of the study area (Subsection 3.3.3.1)
- ecological land classification of the site including dominant cover species (Subsection 3.3.3.1)
- topography and surficial geology of the study site
- hydrological characteristics of the site (e.g., surface water present at the site, drainage characteristics of the soil)
- soil classification of the study site including soil chemistry (Subsections 3.3.3.2 and 3.6.1)
- soil contamination and physical and chemical characteristics at the site

In addition to information about the site, background data also needs to be collected in preparation for sampling and testing activities. Examples include:

- identify ownership of the site (e.g., land titles, legal surveys)
- obtain access permits for the site
- identify what facilities exist on the site that are available for use by the sampling team (e.g., water source for equipment decontamination, on-site laboratory facilities, access to shipping services for samples)
- identify local hospitals and emergency numbers in case of accidents or injuries
- if required, conduct a ground disturbance survey to identify any disturbance of soils at depths > 30 cm in order to locate underground utilities that might constitute a digging hazard (e.g., pipelines) (Hatscan, 2006) — in some jurisdictions formal ground disturbance training might be required.

The sources of these background data include (Mason, 1992; Paetz and Wilke, 2005):

- detailed maps (soil, vegetation, topographical)
- remote sensing imagery (e.g., aerial photographs or satellite imagery)
- reports, publications, and studies generated by surveyor's offices, geological surveys, industrial inspection authorities, mining boards, mining companies, oil and gas companies, geotechnical institutions, regional and municipal archives, local museums, agriculture, geological and forestry authorities
- environmental studies from university and government institutions
- environmental impact or *assessment studies*
- site surveys

Web-based references to Canadian resources (e.g., soil maps, vegetation maps, land cover information) can be found in Appendix F.
In addition to collecting background data arrangements should be made, prior to travel to the site for sampling, for the shipment of soil samples to the testing laboratory(ies).

3.3.3.1 Ecological Classification

Ecological classification of the study area is one of the first steps when conducting field assessments, either at the reconnaissance or site level. Ecological classification is based upon a holistic view of the interactions and links among the following ecosystem components: geology, landform, soil, vegetation, climate, wildlife, water, and anthropogenic activities. Ecosystem processes, their interactions, and the relative dominance of their individual or combined effects within that ecosystem are used to spatially define and classify areas on an ecological basis. Classification of the ecosystem(s) present within a study area provides access to a pool of knowledge on the processes at work within that environment and allows pre-planning of sample locations on a broad scale. It also aids in the identification of an appropriate *reference site*(s) and test species.

The National Ecological Framework for Canada (Marshall and Schut, 1999) defines the ecological land classification and the ecozones and ecoregions of Canada. Other classification systems for provincial/territorial levels and for specific ecological regions (e.g., forested regions) are listed in Appendix B and references

to additional Web-based resources (e.g., mapping) are provided in Appendix F.

3.3.3.2 Soil Surveys

The field description of soils is an important component of a contaminated site assessment and should be prepared prior to, or at the same time as, collecting soils for biological testing. Description of the soil in the field provides a basis for identifying possible contamination or toxicity confounding factors, selection of samples and test species for field screening and testing, and interpretation of test results (AE, 2008).

To describe the soil at a site the investigator can first refer to soil maps of the study area. A soil map shows the distribution of one or more soil types or classes within a physiographic region. The distribution of the soil type(s) on a soil map is a result of the combined influences of the physiographic region (climate and parent materials), vegetative communities (ecoregion and dominant plant species), and landscape position (topography and surface water movement) that occur in that region. To generate a soil map, a soil survey is conducted, usually with a specific land use in mind, and soil map units are developed based on the characteristics of the soil that are significant to that land use.

Soil survey and map information for a given area are available through both federal and provincial sources. For example, soil surveys completed within each province and/or territory are available online or in hard copy from the Canadian Soil Information System (CanSIS) and its National Soil Database (http://sis.agr.gc.ca/cansis/nsdb/intro.html). If formal soil surveys have not been conducted within an area, it is possible that data have been collected/summarized by industries operating within an area and/or university research projects. If soil maps are not available, then it is likely that geological maps (Geological Survey of Canada) exist for that area and can be used as a stepping stone for soil survey/classification within a region. A detailed review of soil survey and mapping can be found within the document "*A Soil Mapping System for Canada: Revised*" (AC, 1981).

3.3.3.3 Site Surveys

An initial site survey should be conducted as part of any contaminated site assessment. In the context of this guidance the survey would either have been conducted as part of the screening level assessment or conducted prior to collecting soil samples for biological testing.[3] Preliminary site surveys are recommended because important information can be obtained to help improve the efficiency and accuracy of the sampling plan and aid in the selection of appropriate biological tests and test species.

- Visual site inspection – provides information such as soil discolouration or the presence of visible contamination (e.g., areas of greatest soil contamination or chemical "hot spots").
- Vegetation surveys – vegetation over- and understory, changes in plant cover or lack of plant cover (e.g., areas of greatest impact or toxicological "hotspots"), or the presence or absence of habitat suitable for ecological receptors of concern can be observed [Department of Environment (UK), 1994] .
- On-site chemical detection equipment – such as X-ray fluorescence or mobile field atomic absorption detectors (to measure metals), colourmetric tests [to measure petroleum hydrocarbons, total polycyclic aromatic hydrocarbons (PAHs), and some explosives] (AE, 2008), and portable gas chromatographs or hand-held photo-ionization detectors (PIDs) (to measure volatile organic contaminants) can be used to refine the sampling plan (EC, 2002a, 2002b).
- Soil salinity field screening tests – conducted with calibrated electrical *conductivity* probes, soil water solutions, and saturated paste extracts.
- Soil pH field screening tests – can be conducted using litmus paper, a calibrated probe, soil extracts, or solutions.

[3] A site survey in the context of this guidance is distinct from a legal survey; a legal survey is conducted to delineate the boundaries of a contaminated area to ensure that the site assessor does not trespass and to identify any movement of contaminant across legal boundaries.

- Soil chloride concentrations – can be estimated using screening tests (e.g., Quantabs™).
- Collection of exploratory samples for biological screening tests – screening tests could include earthworm acute mortality and avoidance tests, microbial contact toxicity tests or functional tests, Phytotoxkit™ tests and seedling emergence and root elongation tests (Table 2).
- Collection of exploratory chemistry samples (if the study is a screening level study and no chemistry samples have already been collected).

Once the survey has been conducted and results have been obtained, a site survey report should be completed and used as part of the suite of background data.

3.3.4 Selection of Sampling Locations

This section and the following outline the general considerations for the selection of soil sampling locations and provides guidance for the selection of soil sampling strategies according to general study objectives and biological test designs.

The selection of sampling locations depends upon the study objectives, the data quality objectives, preliminary information, and on-site conditions (Paetz and Wilke, 2005). Table 7 provides a summary of sampling strategies, including those discussed in the following subsection along with their respective selection criteria; readers are referred to Appendix C for more detailed information and guidance.

Examples of on-site conditions that need to be considered when designing a sampling strategy include:

- local topography (e.g., deposition zones for airborne contaminants; low areas where liquid contaminants from spills could concentrate)
- proximity to water (e.g., riparian zones where water might be the source, or sink, for contaminants)
- climatic conditions (e.g., prevailing winds)
- tree cover
- gound cover
- habitat for ecological receptors
- soil type and/or soil physicochemical characteristics (e.g., pH, *texture*, organic matter content)
- bedrock type
- location of contaminant source (point or non-point)
- direction of contamination (unidirectional; diffuse)
- background concentrations of potential contaminants of concern
- physicochemical properties of the contaminants of concern (e.g., gas, solid, liquid, volatile, water-soluble)

3.3.5 Sampling Strategies

There are many different types of sampling strategies. Two main categories are probabilistic and non-probabilistic sampling. In non-probabilistic (also called targeted, convenience, or judgmental) sampling strategies, sample locations are selected based on expert knowledge of the site or on professional judgment. With probabilistic sampling strategies, sampling locations are selected by applying statistical theory and random chance to location selection (USEPA, 2002a). Non-probabilistic sampling strategies can be more time efficient than probabilistic strategies if knowledge of the contamination at the site is available; however, there is no way to measure the precision of the data, and the data cannot be interpreted statistically. In contrast, probabilistic sampling strategies are more difficult to implement (often requiring the assistance of a statistician), but when used, the uncertainty associated with the data can be measured, quantitative conclusions can be made about the toxicity of the sampled soil (e.g., statistical inferences can be made about the data), and decision error criteria can be addressed (USEPA, 2002a).

Before selecting a sampling strategy, it is important to keep in mind that the choice of sampling strategy is driven primarily by the study objective and secondarily by site characteristics. A clear statement of study objectives is necessary to cost-effectively and defensibly select a sampling strategy.

The choice of sampling strategy depends on:

- the nature of the site or portion thereof (sometimes refered to as a statistical population[4]) and contaminant situation (e.g., type and spatial distribution) being sampled;
- the types of inferences that will be made using the data collected (e.g., the types of questions that need to be answered using the data), and,
- pragmatic considerations such as cost, site accessibility, etc.

If a sampling strategy is chosen that is inconsistent with the first two of these factors, inferences made using the sample may be biased or even incorrect. Example A illustrates this potential.

Example A. Consider assessing an industrial site for petroleum hydrocarbons (PHC) where an environmental decision is driven by an average concentration not exceeding some criterion. Often, the majority of samples will be collected around known or suspected point sources within an industrial site with a lesser number of samples collected in areas not thought to be affected by PHC. A simple mean of these data will bias the overall site mean upward, biasing the environmental management decision. If the purpose of the assessment is to find "hot spots" on the site, collecting the majority of samples at known sources only confirms what is known and does not greatly reduce the probability of missing unknown hot spots. Finally, if the purpose of the sampling is to delineate areas of contamination (for example, to define *isopleths* consistent with an environmental decision criterion such as a specific contaminant concentration), an alternative sampling strategy is warranted.

It is critical that the intended use of the data collected be considered before collecting samples. Dialogue between all stakeholders is recommended to ensure that the goals of the sampling strategy are met in the most cost-effective and scientifically defensible manner.

A sampling strategy is used to collect samples from a site. Therefore, a definition of the site about which inferences will be made is required prior to deciding upon a sampling strategy. For the purposes of biological testing of soil samples, the site is defined as the delineated tract of land that will be characterized biologically. Once the site is defined, discrete elements of the population comprise the samples. Because contaminants within soil, and soil itself, both have a continuous or quasi-continuous distribution over the site, discrete "elements" within the site cannot be selected. Therefore the concept of selecting an element (the sample) from a list of all possible elements (often referred to as a "sampling frame" in the statistical literature) does not apply; rather a location within a site is selected for sampling. The continuum in both soil composition and contaminant distribution within a site, and an inability to select discrete elements within a site affects how a "sample" is collected and even defined. Example B illustrates these ideas.

Example B. A soil sample may be collected using a deep core that will consist of the surficial organic layer, various *soil horizons* and possibly weathered parent material. Contaminant concentrations can vary with depth within this sample as a function of topography, porosity, sequestration characteristics of the contaminant(s), etc. Thus an average concentration from such a "sample" might be very different from a "sample" that contains soil from only one horizon. It is important that the definition of a "sample" within a continuum such as soil reflects not only the concerns of stakeholders but also the contaminant type, relevant exposure pathways, etc.

---

[4] From a statistical point of view, the term "population" is used in sampling program design to denote contamination measures or toxicity screening values within a site.

It is evident from the preceding example that the definition of the term "sample" may be site-specific. Thus in the context of the guidance in this subsection, the term "sample" is used in a generic sense.

3.3.5.1 Error Rate Control

Consider a set of biological measurement endpoints collected from a site at a given time. The set of biological measurement endpoints can be variously summarized; for example, the median might be used to represent the "middle" or "centre" of the dataset. If another set of data were collected at slightly different locations or on a different day, a slightly different median would be produced. Repeated sampling will produce a variety of medians clustered around some central value; the value obtained from a specific collection of data is only one of this possible set.

The conclusion reached following a hypothesis test is a function of the particular median or medians estimated (and by extension, the set of biological responses collected). It is possible (but unlikely) that even if adverse biological effects occur on a site, a particular set of samples will show no or little effect. If that (unlikely) median were compared with say, a reference median for a reference exposure study or to a regulatory decision point, the conclusion might be that the null hypothesis of no biological effects cannot be rejected. This conclusion is incorrect since we "know" that adverse biological effects occur. This error or mistaken conclusion is called a Type II error (usually designated by the Greek letter beta; β) and leads an environmental manager to the incorrect conclusion that there are no adverse biological effects at the study site.

Another type of error that can arise when testing hypotheses is a Type I error (usually designated by the Greek letter alpha; α). This occurs when a particular dataset leads to the incorrect conclusion that there is a significant adverse biological effect when in fact no such effect occurs.

For aquatic receiving environments EC (2002c; 2004b) and INAC (2009) advocate setting the Type I error rate equal to the Type II error rate, thereby reflecting equal risk of an incorrect conclusion being in favour of the environment or those discharging to the environment. However, in the case of special environments (*wetlands*, heritage sites) the Type II error rate might be set lower than the Type I error rate.

In Canada, at this time, there are no specific legislated error rates. Some authors have suggested that either error rate should be no more than 20%. With respect to protection of aquatic receiving environments, EC (2002c; 2004b), discusses error rates of 10% and INAC (2009) recommends that the Type II error rate be $\leq 10\%$.

Given the importance of this topic, error rates should be discussed when a biological assessment study of contaminated soils is designed and the Type I error rate must be specified before data are statistically analyzed otherwise the validity of any conclusions reached is suspect. Type I and II error rates, that define the levels of uncertainty and variability in the data acceptable for the decision maker, represent important examples of data quality objectives that should be determined at the outset of the study, as discussed in Subsection 3.3.1. Topics for discussion of error rates include: 1) the relationship between error rates, the degree of environmental protection, and biological assessment/remediation costs; and, 2) the need to set differential Type I and II error rates. The relationship between Type I and II error rates is summarized in Table 6.

The remainder of this subsection describes some commonly encountered sampling strategies and provides recommendations for use; however, without knowledge of a particular study objective such guidance can only be of a general nature. The description of these sampling strategies is augmented by case studies presented in Appendix C where sample size calculations are also described; readers should also consult Appendix C for more detailed guidance. The strategies described in this subsection were selected on the basis of: applicability to soil toxicity sampling designs; their fundamental nature as a foundation for more complicated

strategy; and, ease of use and data interpretation by non-statisticians. More cost-effective, nuanced (and considerably more complex) sampling plans exist. Some of these are discussed in Cochran (1977), Gilbert (1987), Thompson (1992), and USEPA (2002a, 2002b). The advice of a statistician should be sought for complicated sampling scenarios to ensure that data collected are suitable for the intended purpose. Additionally, a statistician might be consulted to optimize strategies where the cost of collecting information or making an incorrect conclusion is very high or a limited budget requires reconciling various objectives.

3.3.5.2 Simple Random Sampling

In a simple random sampling (SRS) strategy, every possible sample at a site has an equal probability of being chosen. The process of randomization can control for the effects of unknown confounding factors[5] which is the primary strength of SRS. Some form of simple random sample selection is part of all probabilistic sampling strategies.

The advantages of SRS are:

- ease of design since locations on a site are randomly selected
- unbiased estimates of the mean and variance
- ease of data analysis by non-statisticians due to simple formulae and availability of software and necessary tables

The disadvantages of SRS are:

- it does not incorporate knowledge regarding known or unknown confounding factors; the failure to use available knowledge can make this strategy less efficient[6] than other strategies; statistical analyses can incorporate effects of known confounding factors, but only if by chance sufficient information is available within the samples that were randomly collected
- it can by chance provide poor coverage of a site
- it may be difficult to implement on some sites due to access issues

Random sampling strategies can be chosen when all factors affecting soil toxicity are homogeneous across the site. This presumes *a priori* knowledge of contaminant concentrations and test-specific toxicity modifying factors at the site. If there is not a reasonable degree of confidence that these (and possibly other) factors are homogeneous across the site, another sampling strategy should be chosen. A case study (Case Study 1) using a simple random sampling strategy is described in Section C.2 (Appendix C).

3.3.5.3 Stratified Random Sampling

Stratified random sampling (StRS) strategies select samples randomly within a stratum. A stratum may be defined from the perspective of effects on the measured variable (e.g., biological test results). An example would be soil composition, which is a factor that can affect biological test results and may not be homogeneous across a site. Biological test results from samples collected from a clay loam lens will likely be different than toxicity test results from samples collected from sandy loam. If the objective of the study is to understand how toxicity varies across the site it is important to ensure that samples are collected from each of these strata.

[5] A confounding factor or variable is any variable that is not of direct interest but may confound the interpretation of the data. For example, slope can affect the distribution of chemical concentrations across a site and if not recorded can confound interpretation of the data.

[6] Within this document efficiency refers to the number of samples required to achieve a given level of precision for a statistic.

Table 6. Relationships between Type I and II Errors

Decision	Null hypothesis is correct	Null hypothesis is incorrect
Do not reject null hypothesis	Correct inference	Type II error (β)
Reject null hypothesis	Type I error (α)	Correct inference

Another example would be a site with very shallow or stony soil consisting of pockets of deeper soil where a lot of biological activity occurs. The site might be stratified based on these ecologically productive micro-sites.

Strata may also be defined for pragmatic reasons such as ease of sampling, cost of collecting samples in different parts of a site, the need to use different field techniques due to varying site characteristics.

Advantages of StRS are:

- when confounding factors are known or understood, sampling that acknowledges the confounding factors (rather than ignoring them as in simple random sampling) are, for an equivalent amount of sampling, more statistically powerful due to the apportioning of variance among strata, rather than within the residual error; consequently, stratified random sampling strategies may be more efficient and cost effective than simple random sampling;
- a stratified random sampling strategy also allows for testing hypotheses about strata; in a simple random sampling strategy, sufficient samples may not be collected within a stratum to test such an *a posteriori* hypothesis; this is particularly true if a stratum comprises a small spatial fraction of a site;
- the stratum-specific precision required for some statistic (such as a mean) can be specified in advance; this can be helpful in situations when differential variability occurs due to the stratum (e.g., soil fertility); and,
- the overall cost of an experiment may be reduced by limiting sampling within strata that are costly to sample

Disadvantages of StRS are:

- the strata must be known prior to sampling;
- StRS strategies optimized to reduce costs or achieve a pre-specified level of precision may fail to achieve the desired benefits if the strata are poorly defined;
- StRS may bias site-level estimates (e.g., mean toxicity at a site) if the probability of samples falling within a particular stratum are not known; therefore, statistical interpretation of data must consider this probability

A consideration[7] with stratified strategies is that analysis cost increases proportionally with the product of the number of levels in each confounding factor. For example, consider two confounding factors such as organic content and percent sand. If each of these two confounding factors has 2 levels, say "high" and "low" there will be 4 strata to consider. If each of these two confounding factors has 3 levels, say "high," "medium," and "low" there will be 9 strata to consider which more than doubles the sampling effort. Therefore, to characterize three levels in each of two factors, we require adequate[8] sample sizes in each of $3 \times 3 = 9$ cells of the sampling strategy.

[7] The word "consideration" is used here rather than disadvantage because the presence of many factors that can affect toxicity is not a consequence of the sampling scheme but rather the reality of the interactions between biotic and abiotic components of the environment.

[8] Statistical methods such as fractional factorial designs can be used to reduce sampling effort within cells of the sampling design; advice from a statistician should likely be solicited.

Stratified sampling strategies are useful to deal with the effects of known confounding factors. However, the spatial extent of confounding factors may not be equal across the site, and if not considered, a disproportional allocation of sampling effort in one or more of the strata might result. Example C further illustrates this point.

Example C. Consider a site in which the confounding factor, soil matrix, has two levels, sandy and clay loams. Ninety percent of the surficial area is comprised of a sandy loam, and 10% as a clay loam. It is likely, due to sorption mechanisms that the bulk of a contaminant such as PHC will reside in the clay loam with some associated small-scale heterogeneity. In this case, a "better" assessment of soil toxicity would be to disproportionately allocate sampling effort among the two strata, not in the ratio of 9:1 as indicated by the spatial extents of sandy loam to clay loam, but rather in consideration of the greater expected variability in the clay loam strata.

In the example above, considerations such as differential sequestration of the contaminant being investigated and differential variability among levels of the confounding factor (soil matrix) lead to sampling efforts that are disproportional among the levels of the confounding factor, soil matrix. When factors that suggest disproportionate sampling are quantitative such as cost of sampling or variability, statistical tools are available to optimize the experiment according to the intended use of the data. A case study (Case Study 2) using a stratified random sample strategy is described in Section C.3 (Appendix C).

3.3.5.4 Systematic Sampling

In systematic (Sy) sampling strategies, a single location is randomly selected and then the remaining locations are sampled at specified intervals. The simplest case of systematic sampling is a transect. In the two-dimensional case such as choosing surficial soil sampling locations at a suspected contaminated site, a grid is randomly overlaid on the site (or site map) and grid intersections become the sampling locations. The grid may be constructed such that the desired number of sampling locations is obtained, or more typically, a given number of grid intersections are skipped to ensure that the desired total number of samples is collected.

One concern with this approach is that if the grid interval matches a periodicity in the environmental sample, inferences from the data may be biased. Various soil remediation efforts could conceivably induce a pattern or periodicity in a site being assessed for soil toxicity. For example, remedial efforts such as injection of bacterial cultures, or physical introduction of materials through mechanical means such as chisel ploughing, could induce a spatial pattern over a site. If the induced pattern matches or is similar to the overlaid grid, there may be bias in the results. This concern, although unlikely, should be considered when sampling soils to assess toxicity.

The advantages of systematic sampling are:

- can be easy to implement
- do not require *a priori* knowledge
- ensures good spatial coverage over the site (but should not ignore known confounding factors)
- easy to use in a sequential manner
- can produce unbiased estimates of the mean [assuming the formulae and systematic sampling strategies variants described in Section C.4 (Appendix C) are used]

The disadvantages of systematic sampling are:

- ability to assess effects of known confounding factors may be compromised
- periodicities in the field that are aligned with the grid can bias estimated statistics
- variances are not as easily estimated as in other designs. In particular, when sampling contaminated soils, the correlation among observations induced by dispersal mechanisms (e.g., wind) will bias variance estimates, almost certainly upwards. A variety of

approaches for estimating variances of data collected under systematic sampling are discussed in Bellhouse and Sutradhar (1988) and Iachan (1982).

Systematic sampling along a transect is not likely the best use of resources when using toxicity tests to estimate a threshold in toxicity. If possible it is best to collect some samples at the "top" (highest concentration), and "bottom" (lowest concentration) of the gradient with the remainder of the samples in the vicinity of the suspected threshold.

Systematic grid sampling is recommended when: a site is suspected to be heterogeneous and there is little background information known to guide sampling; correlation among adjacent biological tests results is likely; it is desirable to delineate toxicity; and, it is desirable to understand how toxicity varies over the site, to find a "hot spot,"[9] or to estimate the probability of missing a "hot spot." Systematic grid sampling is also recommended when the data are to be used for geostatistical modelling to generate biological response maps of the site. An adaptation of systematic sampling is presented in Case Study 3 (a 1-dimensional transect study) (Section C.4; Appendix C). Section D.2 (Appendix D) describes how geostatistical methods are used with data derived from systematic sampling designs to generate different contour toxicity maps. It is recommended for readers without a background in *geostatistics* first become familiar with the material in Subsection 3.3.6 and Section D.1 (Appendix D) before reading the case study in Section D.2.

3.3.5.5 Adaptive Cluster Sampling (ACS)

Adaptive cluster sampling (ACS) is a two-stage sampling strategy. In the first stage a probabilistic sampling strategy is used to sample the site. This might be a simple random sample, a stratified sample, etc. Then at selected locations of particular interest additional samples are collected following some decision rule.

This type of sampling program can be very cost-effective if initial sampling can be based upon a low-cost measurement. This measurement might be a field-based measurement or a laboratory-based measurement with a short turn-around time.

Advantages of ACS are:

- can be cost effective
- is useful for delimiting areas of interest

Disadvantages of ACS are:

- there is no control over the total sample size because it is not known in advance how many discrete locations will exceed the criterion for cluster sampling
- special techniques must be used to estimate statistics such as means
- the time interval between first and second rounds of sampling may be unacceptable

Adaptive cluster sampling is appropriate to use when it is known that hot spots exist but there is some uncertainty about the number or size of the hot spots and a rapid or relatively inexpensive surrogate measurement is available for the initial probabilistic sampling strategy. Some adaptations of systematic sampling are presented in Case Study C.4 (Appendix C).

3.3.5.6 Ranked Set Sampling

Ranked set sampling (RSS), like adaptive cluster sampling, combines two stages of sampling. There are various forms of RSS; only the simplest RSS with simple random sampling as applied to soils is discussed here.[10] In the first stage of sampling, m^2 sample locations within a site are randomly selected. Here, "m" is the number of ranks (usually a small number such as 2 to 7). Then, the m^2 locations are randomly allocated to m "sets." Within a given set the locations are ranked according to the variable of interest using either professional judgment or information from a rapid or screening level test

[9] Hot spots are further discussed as Case Study 4 in Section C.5 (Appendix C).

[10] Other forms of RSS randomly select sample locations within strata, along transects, etc.

(laboratory or field test) that is strongly correlated with the definitive biological test results. Then the smallest (or largest) ranked sample location is selected for the definitive laboratory biological test(s) of interest. This procedure is repeated with the next set of locations until a complete set of ranked samples has been collected for definitive biological testing.

For example, if the ranking procedure enables discrimination at three levels, "low," "medium," and "high" m = 3, then 9 locations are randomly selected at a site. Three of these locations are randomly assigned to set "A," three to set "B," three to set "C." The locations within set A are ranked as "low," "medium," or "high" using the rapid screening test and the sample from the location designated as "low" is selected for further biological testing. This procedure is repeated for set B and the location designated as "medium" is selected for biological testing. The procedure is applied to set "C" and the location designated as "high" is selected for biological testing. Thus, in this example, three samples are obtained for biological testing[11]. The entire procedure is repeated r times to conduct $n = r \times m$ biological tests. Note that when ranked set sampling is used, it is very important that the screening-level tests (or professional judgement) are strongly correlated with the definitive biological test results.

Advantages of RSS are:

- the standard error of the mean is lower than that of an equivalent simple random sampling strategy (McIntyre, 1952)
- can be very cost effective if the collection of the ranking information is much less costly than that of the laboratory test
- ensures that the overall conclusion is not unduly affected by unusual observations

Disadvantages of RSS are:

- the ranking methodology (either expert judgment or in the case of soil biological testing more likely a field test) is accurate with respect to the true ranks
- data interpretation is more complex than when using simple random sampling

Ranked set sampling should be used when an accurate ranking method is less costly than the biological test method(s) employed to make inferences.

3.3.5.7 Available Site Knowledge

The ability to choose the most effective biological sampling plan for a given purpose depends upon the available site knowledge. This may range from the most general knowledge regarding site topography to detailed knowledge regarding contaminants of potential concern, measurement response-specific toxicity modifying factors and screening or even definitive level biological test results. These ideas are presented in Figure 1. Background data are collected early in the study plan stage, in order to inform the selection of the sampling strategy and the development of the sampling plan (Section 3.3.3).

The selection of any particular sampling strategy or combinations of different strategies is site-specific and depends upon the study objectives and characteristics of the site. Regardless of the sampling strategy selection, available site knowledge should be used in its implementation.

3.3.5.8 Representative Samples

Regardless of the type of sampling strategy used to collect soil samples for biological testing it is important that the samples collected are representative of the site that is under assessment. A sample is considered representative if it accurately and precisely

[11] Note that in one aspect randomness is omitted in that the rank of the toxicity sample is deliberately selected (e.g., low, medium, or high) within a set. However, each element of the set is randomly selected within that set using rank information. Some entire sets will by chance produce higher estimates of toxicity even for the "low" within-set sample than for another set. When this is repeated over sets, a good coverage of the distribution of toxicity is obtained across the site.

represents the characteristic(s) of the soil type, the location, or the environmental conditions at the site (USEPA, 2002a). This means that if replicate samples collected from a site are representative, then any variation among the replicates will be due to random variation of the characteristics of the site.

Collecting a representative soil sample can be a challenge due to the inherent heterogeneous nature of soil characteristics and, distribution of contaminants. The spatial heterogeneity of soil (> 50% of the variability between similar soils occurs within 1 m) is well-documented. In general, larger soil samples are more representative than smaller samples due to the inherent heterogeneity of soil (CCME, 1993a). As *soil properties* can strongly influence the bioavailability of contaminants and performance of test organisms, this heterogeneity can significantly influence the results and precision of biological test results.

To collect representative samples, it is necessary to have a clear understanding (good site knowledge) of the contaminated site of interest (the statistical target population). In order to decide whether a sample might be representative before collecting samples it is necessary to list all factors that can affect the biological response being measured and determine over what area within the site the set of factors is "homogenous." The definition of homogenous is subjective and contextual. For a toxicity test response a "homogenous" soil area would produce not much more variation in the response than 1) allowed for as a test acceptability criterion; or, 2) in a suitably matched reference soil. Soil samples randomly collected within such a homogenous area of soil (and that follow suitable physical sampling protocols) should be "representative" samples. Different sampling strategies can be used to collect representative

Environmental Fate and Transport[3]	Contaminant Concentration and Location	Toxicity Modifying Factors[1]	Biological Toxicity	
surficial topography	none	none	none	
	site usage history/anecdotal		some screening level[2] tests	
limited knowledge of surficial porosity, TOC[4], particle size distribution, etc.	limited judgmental sampling data	limited geospatial information	⬇	low
			quality screening level[2] test isopleths	moderate
geospatial knowledge of surficial porosity, TOC[4], particle size distribution, etc.	quality isopleths / pre-defined areas of interest	quality isopleths / pre-defined areas of interest	some definitive test results	high

Increasing Knowledge

[1] presumes relationship between contaminant-specific toxicity modifying factors (TMFs) and measurement endpoint is known

[2] requires demonstrable correlation with definitive biological assessment measurement endpoint

[3] unnecessary if contaminants of potential concern geospatial data available

[4] total organic carbon

Figure 1. Relationship between site knowledge and background data

samples when there are two or more "homogeneous" areas within a site (e.g., stratified random sampling strategy).

However there is a possibility that unknown factors are present within a "homogenous" sample area in a site such that data exhibit characteristics of a mixture of statistical populations. In this case, the samples are not "representative." Guidance on identifying mixtures of statistical populations is beyond the scope of this document. Some very general "rules-of-thumb" are presented below and readers are encouraged to consult with a statistician if there is concern that a set of samples seems to "represent" more than one population.

Rules-of-thumb for identifying multiple populations:

- If the *coefficient of variation* for a biological test response is > 100% it is likely that the data arise from two distinct groups or populations.
- Abrupt changes in the empirical cumulative distribution function[12] usually indicate distinct populations.
- If in a data set of a given biological endpoint there are observations outside the range defined by the mean ± 3 standard deviations this may indicate distinct populations.

It is important to note than an otherwise "representative" sample may not represent the site if the physical sampling methods or analytical/inferential methods induce bias and/or imprecision. As a result, guidance on appropriate physical sampling methods for the purposes of biological assessment is provided in Sections 3.6 and 6.1.1. Inclusion of split sample (Subsection 3.3.11.2) collection and analysis in the study and sampling plans is recommended to ensure the representativeness of samples collected. However, note that a satisfactory split sample analysis would not exclude the samples from being "unrepresentative" on the basis of site-related factors.

Soil toxicity testing protocols (e.g., EC 2004a, 2005a, 2007a) provide guidance on how to generate a biological response of a soil sample that controls for precision in an unbiased manner.

3.3.6 Use of Geostatistics in Sampling Contaminated Soil for Biological Testing

Geostatistical methods can be used to develop spatial models that can:

a) generate stochastic contour maps of observed biological responses at a contaminated site (soil toxicity maps);

b) make predictions of biological response for unsampled locations at a site; and,

c) can test hypotheses that were developed in response to specific study objectives.[13]

Geostatistical models are particularly useful in that they acknowledge variability, and allow for hypothesis testing and prediction with confidence intervals. Geostatistical tools may be used in the development of sampling strategies and to analyze the data generated from the soil sampled (Mason, 1992).

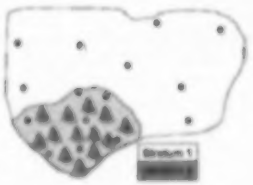
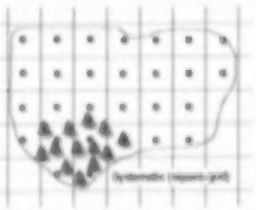
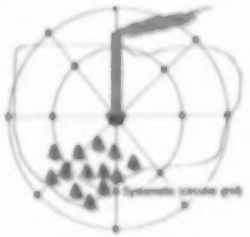
---

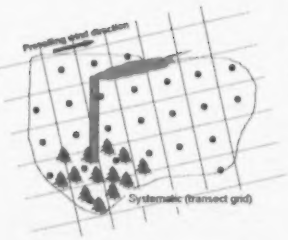
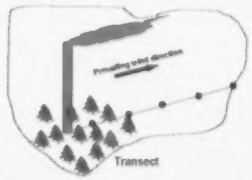
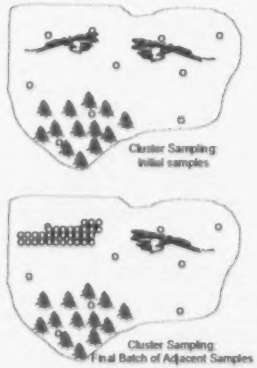
[12] An empirical probability *density* function may be thought of as a smoothed frequency histogram of the responses. The familiar bell-shaped normal distribution is an example of a density function. The density function can be used to provide estimates of probabilities; if these probabilities are cumulated they sum to one and if plotted a *cumulative distribution* function is produced. The cumulative distribution function for the normal distribution a sigmoidal curve with a horizontal asymptote at 1. Most software packages can produce an empirical (empirical means the data are used to generate the distribution function rather than a theoretical function) cumulative distribution function.

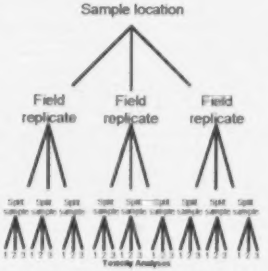
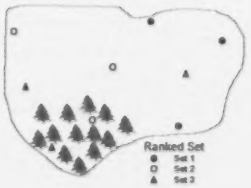
[13] Note that commonly encountered hypotheses such as comparison of means between a reference and a contaminated site generally use an experimental design-based approach to collecting samples. The experimental design (e.g., comparing two means) strongly influences the type of sampling strategy selected to collect soil samples. Examples of different sampling strategies and the various experimental designs they support are discussed in Subsection 3.3.5.

Table 7. Sampling strategies and selection criteria for the toxicity assessment of contaminated soil

Strategy	Study objective(s)	Selection criteria/Site knowledge	Strategy description		Comments	Reference(s)
Non-probabilistic						
Opportunistic sampling	Preliminary investigation; limited access or time	Access to site difficult; limited time	Selection of locations has no predetermined pattern		Should be conducted by an experienced investigator; not recommended	Paetz and Wilke, 2005
Targeted	Site screening; emergency sampling; hot-spot sampling	Constrained budget or schedule; reliable site information exists; site relatively small; small # samples to collect	Selection of locations based on professional judgment and/or historical information		Level of sampling uncertainty cannot be measured; statistical inferences not possible	USEPA, 2002a; Mason, 1992; Paetz and Wilke, 2005; ISO, 2002a, 2003a, 2005b, 2006d
Probabilistic						
Simple random sampling	Site screening; average toxicity at a site; compare site to reference or regulatory criteria; remediation evaluation	Site is relatively uniform; little site information available	Sampling locations selected so that each sample has the same chance of being taken from any given location		Simplest but least efficient strategy; basis for many other strategies; protects against bias; sample locations may not be spread evenly over site	Cochran, 1977; Thompson, 1992; USEPA, 2002a; Mason, 1992; Paetz and Wilke, 2005; ISO, 2002a, 2003a, 2005b, 2006d

Strategy	Study objective(s)	Selection criteria/Site knowledge	Strategy description	Comments	Reference(s)	
Stratified random sampling	Average toxicity at a site; sample along concentration gradient (concentrations can be strata); compare site to reference or regulatory criteria; determine influence of soil/habitat type on toxicity; remediation evaluation	Strata on site are well-defined (moderate to high site knowledge); variability within strata expected to be lower than variability among strata	Samples collected from within selected, discrete areas (strata) on site (e.g., soil type, topography, vegetative cover). Samples can be collected within strata randomly or systematically		More efficient and precise relative to simple random sampling; samples more representative; samples provide more information especially if soil variables are correlated with toxicity and/or bioavailability; information can be obtained on risk to ecoreceptors according to habitat	Cochran, 1977; Thompson, 1992; USEPA, 2002a; Mason, 1992; Pactz and Wilke, 2005; ISO, 2002a, 2003a, 2005b, 2006d
Systematic grid sampling	Site screening; identify toxicity hot spots; generate map of soil toxicity (delineation of areas according to levels of biological effects); average toxicity at a site; compare site to reference or regulatory criteria; remediation evaluation; long-term monitoring	No information about site; identifying toxicity on site; identify pattern/extent of suspected toxicity on site (moderate to high prior knowledge of site); map of soil toxicity is end goal	Samples collected in a regular grid pattern over entire site; starting location and grid orientation randomly chosen; grids can be square, triangular, circular, etc.	 	Practical and convenient; precise; uniform coverage; good starting strategy for use of geostatistics like kriging to develop toxicity maps; care must be taken to use appropriate formulae	Inchan, 1982; Hoffhouse and Sutradhar, 1988; USEPA, 2002a; Mason, 1992; Pactz and Wilke, 2005; ISO, 2002a, 2003a, 2005b, 2006b

Strategy	Study objective(s)	Selection criteria/Site knowledge	Strategy description	Comments	Reference(s)	
	Sample along concentration gradient; compare site to reference; generate map of toxicity (delineation of areas according to levels of biological effects); long-term monitoring	Moderate to high prior knowledge of the site; sampling along contaminant plume desired	Samples collected in a regular pattern with one grid axis parallel to contaminant plume axis		Practical and convenient; precise; good for regression-design toxicity evaluations; can use geostatistics to develop toxicity maps	USEPA, 2002a; Mason, 1992
Systematic 1-dimensional transect sampling	Sample along concentration gradient; compare site to reference; long-term monitoring	Moderate to high prior knowledge of the site; sampling along contaminant plume desired	Samples collected in a regular pattern parallel to the contaminant plume axis		Practical and convenient; cost-effective; good for regression-design toxicity evaluations	ISO, 2006c; Spurgeon *et al.*, 2004
Adaptive cluster sampling	Hot-spot sampling; site screening; to delineate areas according to levels of biological effects	Little prior information about site; some information indicating areas or gradients of contaminants of potential concern or toxicity modifying factors	Probabilistically choose initial samples; then collect adjacent samples following predetermined selection rules		Can be expensive if definitive toxicity tests are used and/or if test turnaround time is slow	Thompson, 1992; USEPA, 2002a

Strategy	Study objective(s)	Selection criteria/Site knowledge	Strategy description	Comments	Reference(s)	
Nested sampling	Can be used with any other sampling strategy or study objective	Use if want to identify sources of variability (e.g., location, sample replication, preparation, testing)	Multiple sample taken at each location (field replicates); each field replicate subdivided (split samples); each split sample tested as one lab sample; each lab sample subdivided into laboratory replicates		Provides information on the components of variance in a study; laboratory replicates for toxicity testing can be taken from any level depending on the study objectives and budget	Mason, 1992
Probabilistic + Non-probabilistic						
Ranked set sampling	Average toxicity at a site; compare site to reference site or regulatory criteria	Use when cost of analyses is high relative to an accurate ranking method; some information indicating areas or gradients of contaminants of potential concern or toxicity modifying factors	A set of ranked samples that spans the range of site conditions is used to evaluate toxicity.		More precise than random sampling with same number of samples; cost of ranking samples in field may be high	USEPA, 2002a

Geostatistical data are data that are collected at discrete locations in space. These locations are spatially continuous. Examples of geostatistical data are concentrations of a contaminant in soil, or a toxicity test response corresponding to a soil sample from a specific location, etc. The measurement at a location may be correlated with observations from nearby locations. These are often referred to as small-scale patterns. The observations may also vary over a large scale. These large-scale patterns are often described as trends. Geostatistical data analyses accommodate both scales of variability (Matheron, 1963). The simplest geostatistical models (or equations) describe patterns among observations, simultaneously dealing with small- (correlations) and large-scale variability (trends). The geostatistical model describes the observed patterns as a function of location and possibly other independent variables. Geostatistical tools are described in Ripley (1981), Isaaks and Srivastava (1989), and Cressie (1993).

Biological test responses separated by "small" distances may be more similar than biological test responses separated by "large" distances due to similarity in soil properties and levels of contamination among adjacent sampling locations. It is primarily this correlation that challenges experimental design-based analyses[14]. However, an interest in "describing" correlated observations over an area via modelling is generally the reason for geostatistical modelling. Thus, once a sampling strategy has been selected and data are collected geostatistical modelling begins with an examination of the correlation between observations (measurements of a spatial variable e.g., soil contamination, toxicity) separated by space, with an eventual goal of predicting a biological response at a location where no measurements are taken.

3.3.6.1 Correlation and Variograms

An understanding of correlation is basic to the understanding of geostatistics. The following is a discussion on correlation as it relates to the fundamentals of geostatistics, the *variogram*.

Imagine samples collected using a grid. Estimate the correlation between all pairs of observations separated by one grid unit. Then estimate the correlation between all pairs of observations separated by two grid units, and so on. Plot the magnitude of the correlation on the y-axis, and the "grid units" (or distance) on the x-axis of a graph. This generates an (empirical) correlogram such as the hypothetical correlogram presented in Figure 2. The correlogram (or hypothetical variogram) in Figure 2 shows that the correlation among observations separated by one "grid unit" is very high. However the correlation among observations separated by 4 or more grid units is very low and becomes effectively zero for observations separated by 6 or more "grid units."

Rather than use correlations, for historic reasons, geostatisticians use the covariance among observations to construct similar plots called variograms. A variogram describes the correlation among observations separated by a given distance. Variograms are used to assess assumptions about the data generated at the site and to better understand the (geospatial) biological response at the site. An empirical variogram can indicate:

- whether spatial trends in the biological response are present at the site;
- to what degree observations are correlated within a given distance (this information is useful when site assessments are conducted using hypotheses that assume observations are not correlated (independent); and,
- the heterogeneity of the biological response across the site.

Choosing and fitting a variogram is the first step in constructing a model that describes the variable of interest (e.g., a biological response) in a spatial or geo-referenced context. A more detailed discussion of constructing and choosing

[14] A common assumption of these analyses is that observations are independent of one another.

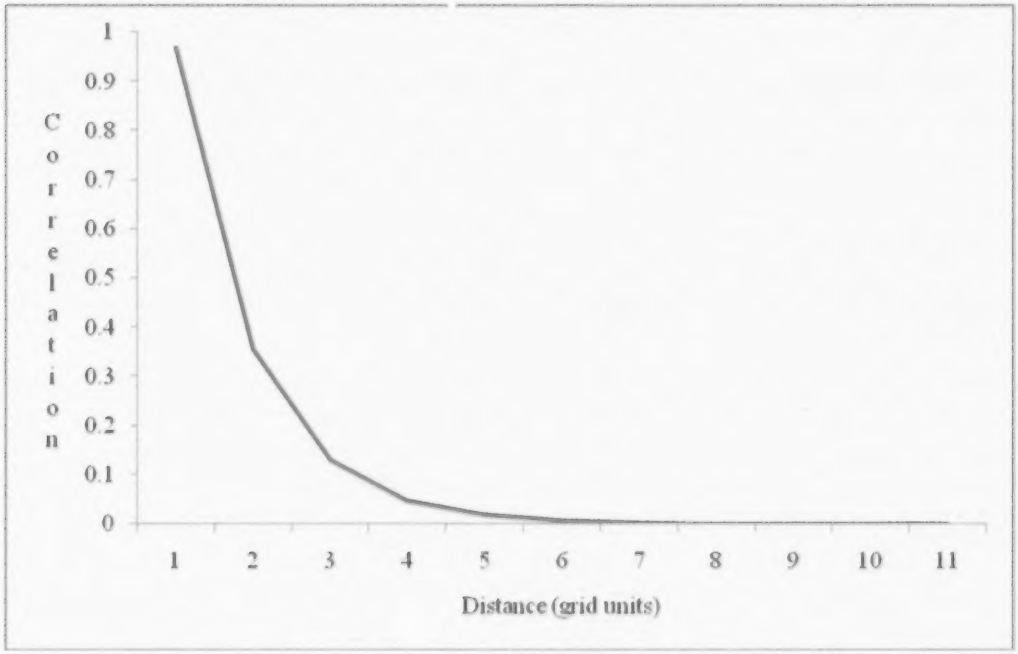


Figure 2. Hypothetical variogram

variograms is provided in Appendix D. This appendix should be referred to by any reader interested in using geostatistics as part of their site assessment.

3.3.6.2 Kriging

Once the correlation among adjacent observations is modelled through the use of variograms, an interpolation method called kriging is used to predict the response of interest (e.g., biological response) at a location where soil samples were not taken and therefore tests were not conducted. The method is considerably more sophisticated than purely mathematical methods, which also generate predictions at unmeasured observations (one common method is inverse distance weighting), because it incorporates the covariance structure of the random function (the variogram) and is "best" because it minimizes the average sum of squared residuals (the mean square error)[15].

Kriging interpolates between nearby observations by ascribing weights to measured observations; some observations have more influence (weight) on a predicted value than other observations. Kriging is the process of generating the weights to minimize the mean square error while acknowledging the correlation structure among observations.

Ordinary kriging assumes that the mean of the biological test response is constant over the area sampled and the correlation between the measurements at given locations can be described by a correlation "model" (the variogram). If there is a trend in the biological responses the mean of the biological responses varies over the area sampled. In this case ordinary kriging should not be used.

Universal kriging may be used when trends are present. Trends are fit using polynomials and thus linear, quadratic, cubic, and higher

[15] For the more statistically inclined, kriging estimates are best linear unbiased estimates.

order[16] trends can be accommodated. Universal kriging can simultaneously estimate parameters for the trend model and the weights described in the context of ordinary kriging.

When estimates of an average response within a specific area or "block" is required, block kriging may be appropriate. Mason (1992) suggests that block kriging may be appropriate when the block is a "remediation unit."

3.3.6.3 Selection of Sampling Strategies for Geostatistical Analysis

If geostatistical analysis is to be applied to the site, the choice of sampling strategies to use is driven by what is being estimated, how it will be estimated, and finally the criterion used to assess the strategy's optimality. Statistical criteria for the selection of the optimal sampling strategy are usually closeness to an unknown parameter, whereas cost criteria may be minimum cost or a ceiling on cost.

When sampling a contaminated soil site in order to obtain the minimum mean or maximum kriged variation[17] across a site, a systematic hexagonal grid is the optimal choice (Olea, 1984). A hexagonal grid is appropriate to use if the function describing correlation among observations is not influenced by direction. This assumption may not be appropriate for contaminated soils in which factors such as slope, soil texture, etc. can induce directional correlations. Therefore a hexagonal grid is not always the optimal grid pattern to use.

Yfantis *et al.* (1987) investigated grid shapes under a different correlation structure[18] than Olea (1984) and concluded [based on the same criteria used by Olea (1984)] that the equilateral triangular design was slightly more efficient than a rectangular grid but that a hexagonal grid was more efficient than either when micro-scale variability is large relative to overall variability and the distance between sample points approaches the distance where observations are functionally uncorrelated.

McArthur (1987) simulated a two-dimensional Gaussian dispersed pollutant and concluded that stratified random sampling plans were the best plans to use for estimating the mean of a locally concentrated pollutant. Strata are defined as concentric rings with a finer grid used in areas with higher pollution concentrations. Implementation of these sampling plans requires prior information regarding the location and pattern of a contaminant.

3.3.6.4 Using Geostatistical Tools to Design Sampling Strategies

Geostatistical tools can also be used in designing sampling programs. For example, a two-stage sampling program is typically used with geostatistical sampling strategies, comprising an initial survey to collect basic data using a radial grid (systematic sampling). These data are used to generate the variogram, which defines the distance over which samples are representative, and the orientation of the correlation structure of a pollution plume. The variogram is then used to dictate the shape, size, and orientation of another systematic grid that will be used in the second, definitive sampling event. Point, or preferably block, kriging is then applied to the data to interpolate between sampled points to make estimates for every block of the sampled area (Mason, 1992).

Geostatistical tools such as kriging can be used to generate stochastic contour maps of soil contaminants, soil properties, or soil toxicity (Sections D.1 and D.2, Appendix D; Fränzle and Kuhnt, 1994; Thomas *et al.*, 1986). Soil contamination contour maps can be very useful to help direct sampling efforts for a toxicity assessment. However, professional judgment and field observations might be just as valuable a tool when designing a sampling strategy (e.g., areas of no plant growth could indicate greater bioavailability of contaminants that would not necessarily correspond with areas of high soil

---

[16] Be wary of over fitting a trend using higher order polynomials. Polynomials of orders higher than three are rarely used to describe trends.

[17] Note that a prediction such as a toxicity test result at any location using kriging has an associated variance. The mean of all these predictions, or the mean variance over the predicted surface, is the criterion being referred to.

[18] Olea (1984) used a linear semi-variogram whereas Yfantis *et al.* (1987) used a spherical variogram.

contamination identified on a map generated by geostatistics) and should be used in conjunction with geostatistics.

Advantages of soil toxicity stochastic contour maps generated using geostatistics are:

- can be useful to help direct remediation activities or to refine exposure and risk determinations of a site-specific risk assessment, particularly when the site is contaminated with a known or unknown mixture of compounds;
- may be used to generate point or block estimates;
- correctly deal with spatial correlation;
- correctly deal with targeted or convenience sampling; and
- could[19] be used in lieu of stratified sampling if the distance between observations is less than the stratum dimensions.

The disadvantage of soil toxicity stochastic contour maps generated using geostatistics is that:

- a large number of samples is often required (ISO, 2005b).

Geostatistical procedures such as those described are useful tools for defining the spatial variability of contamination, though assistance from a skilled geostatistician is recommended to ensure that the data generated are valid (Mason, 1992). As mentioned earlier, geostatistical techniques are discussed in more detail in Ripley (1981), Isaaks and Srivastava (1989), and Cressie (1993).

Readers interested in applying geostatistical methods in their site assessment are first directed to Section D.1 (Appendix D) for further description on how to construct, fit and choose a variogram. Readers are then directed to Section D.2 (Appendix D) if they are interested in geostatistic case studies that provide examples of how to generate and use a kriged surface to interpret biological response data from a contaminated site.

3.3.7 Waste Pile Sampling

The guidance for sampling soils for biological testing provided in subsection 3.3.5 also generally applies to sampling soils from waste piles. Therefore guidance throughout Subsection 3.3.5 should be referred to when sampling waste piles; however, there are some additional considerations unique to waste pile sampling. These include:

- obtaining information about the history of the waste pile [e.g., the process that generated the waste, age of the pile, waste management methods, waste pile heterogeneity, etc (ASTM, 2006c; ISO, 2007c)];
- knowledge of the physical attributes of the waste pile including size, shape, soil physical characteristics (e.g., particle size distribution, *moisture content*), access and stability (compactness) (ASTM, 2006c; ISO, 2007c; US Navy, 2009);
- the chemical stability of the waste pile (NJDEP, 2005; ASTM, 2006c; US Navy, 2009);
- interactions between heterogeneous portions of a waste pile and chemical stability (NJDEP, 2005; US Navy, 2009);
- applicable jurisdictional and regulatory requirements for specific number of samples per waste pile volume (NJDEP, 2005; ASTM, 2006c; US Navy, 2009); and,
- depending on the study objectives, it may be desirable or necessary to sample the surrounding environment to assess potential effects of waste pile storage (ASTM, 2006c).

As with sampling soil from contaminated sites, when sampling waste piles it is important to carefully define the study objectives (Section 3.2; ASTM, 2006c; ISO, 2007c; US Navy, 2009) and data quality objectives, including acceptable

[19] The ability of kriging to replace stratified sampling is a function of the project objective. Stratified sampling programs are often designed such that the within-stratum samples sizes are sufficiently large to detect a change of interest.

levels of decision error (Subsections 3.3.1 and 3.3.5, respectively; ASTM 2006) early in the study plan stage and well before sampling commences.

Prescriptive guidance for sampling strategies for collecting soils from waste piles is not provided in this subsection because site-specific characteristics of any individual waste pile strongly influence the sampling strategy most appropriate for use. However, as with contaminated sites, probabilistic sampling strategies are recommended when sampling waste piles (ASTM, 2006c; ISO, 2007c; US Navy, 2009) in order to collect representative samples and to limit and quantify uncertainty (Subsection 3.3.5). Although some reference documents provide general guidance on the use of some probabilitistic strategies [e.g., simple random, stratified random, systematic grid (ASTM, 2006c; ISO, 2007c; USEPA, 2002a; 2006), ranked set, sequential (USEPA, 2002a; 2006) and adaptive cluster sampling (USEPA, 2006)] all recognize that prescriptive sampling guidance for sampling all waste piles is not possible.

Regardless of the sampling strategy selected, the sampling process might be hierarchical. In the absence of "sufficient" information to characterize a waste pile or to verify a characterization, a first round of sampling for analysis of the waste might be warrented. If the biological response to the samples collected exceed an action level or pre-determined criteria, additional evaluation may be required, depending on the study objectives (NJDEP, 2005).

Any sampling strategy used must consider the study objectives, which might be biological assessment in support of remediation, risk assessment or re-use (ISO, 2007c) as well as desired level of rigour and cost. If more than one analyte[20] is of interest then the sampling strategy should be driven by those analytes of most interest. If the requirements for defensible sampling for one analyte are incompatible with sampling requirements for another analyte, then a separate sampling strategy for each analyte might be required (ISO, 2007c).

A consideration specific to waste pile sampling is the volume of a waste pile that can be represented by one sample. Each waste pile is unique due to the heterogeneity in a waste pile. There are many reasons why a waste pile can be heterogeneous (e.g., how waste pile was constructed, source of waste, reason for creation of the waste pile, etc.) and this heterogeneity has led to the proactive principle that a sample can only represent a limited volume of a waste pile. Depending on the jurisdiction, there can be regulatory requirements for the number of samples per volume of homogenous waste for sampling soil for chemical characterization (BC MELP, 1995; NJDEP, 2005). There is no known requirement for number of samples collected per volume waste pile for biological assessment at this time.

A closely related consideration in waste pile sampling is that of obtaining a representative sample. This again is due to the heterogeneity commonly found within and among waste piles. As described in Subsection 3.3.5, it is important that samples collected for biological testing are representative of the statistical target population of interest, in this case the waste pile that is under assessment. Generally the level of sampling effort is affected by the extent of the heterogeneity within a waste pile (ASTM, 2006c). Although it would appear that random sampling from a waste pile would comprise a representative sample, various authorities on waste pile sampling (ASTM, 2006c ; BC MELP, 1995; ISO, 2007c; NJ DEP, 2005; USEPA, 2002a; and US Navy, 2009) uniformly do not prescribe any specific sampling strategy due to the uniqueness of each waste pile. The guidance provided in Subsection 3.3.5 on collecting representative samples is applicable to waste piles, including considerations when it is suspected that samples have been collected that are not representative. As with collecting representative samples from contaminated sites, the collection and analysis of split samples is recommended (Subsection 3.3.11.2).

[20] A toxicity test measurement endpoint *or* the toxicity of soil containing a specific chemical analyte of interest.

3.3.8 Identification of Sampling Locations

Once a sampling strategy has been chosen, it is important to accurately locate and identify the sampling locations. A common method used to identify the spatial coordinates of a sample location is through the use of the Global Positioning System (GPS) (either universal transverse Mercator (UTM) easting and northing coordinates or latitude/longitude coordinates; both with ± 1 m accuracy); an alpha/numeric grid or distances from landmarks can also be used (USEPA, 2006). Sample locations can then be marked on topographical maps. Contour maps of soil concentrations and/or toxicity are also used at this juncture, if available.

Once at the site, GPS units (widely available from field-sampling suppliers) with the sample location coordinates downloaded can be used to guide researchers to each sample location. Maps marked with sample locations should also be used for confirmation of sample location; GPS coordinates can be conveniently electronically mapped onto one or more types of maps. At the site, the sample locations should be identified with flags, coloured stakes, or other markers prior to sampling. Photographs should be taken before, during, and after sampling, and a sketch map with all relevant information should also be made. Both photographs and sketches should include a scale and direction marker. The horizontal and vertical location of the sample location should be recorded and accurate surveying completed *after* the sample has been taken (Paetz and Wilke, 2005).

If a sampling location needs to be changed due to an unforeseen obstruction (e.g., large stones, a tree, or debris) then the sample location should be changed in the field using contingency plans (contained in the sampling plan, Section 3.4) that were made in advance. Without these pre-determined contingency plans *ad hoc* decisions must be made in the field, and this can lead to bias (ISO, 2002a). If a sample location needs to be re-located, the new location, and the reason for it, should be documented in the sampling report (ISO, 2006d).

3.3.9 Sample Size

The minimum volume or mass of soil required for testing depends upon the study objective, site conditions, and the tests to be conducted. The ecotoxicity laboratory(ies) that is to conduct the toxicity testing should be consulted early in the planning process, ideally when designing the study and during sampling. Environment Canada test methods (EC, 2004a, 2005a, 2007a) can provide descriptions of generic test designs; however, the laboratory that will be conducting the testing should still be consulted as it can provide test designs that meet the specific needs of the study (Table 8).

The analytical laboratory(ies) that will conduct any chemical analyses of the soils collected for testing should also be consulted at the same time, in order to ensure that sufficient soil is collected. The amount of soil to collect depends on the number and type of tests included in the test battery, and:

a) The experimental design of the toxicity tests:

- The replicate requirement for a *treatment* (soil sample) is very different if the test is run with soil samples as individual treatments [e.g., one or more samples and a reference soil(s))] compared with a multi-concentration regression design (e.g., one contaminated sample diluted with a reference soil)
- If the contaminated soil samples are to be diluted with a reference soil, then the volume collected per contaminated soil sample can be significantly greater than that required for one treatment. The amount of reference soil required for this type of experimental design also increases.
- The amount of contaminated and reference soil required for a dilution test depends on the number of dilution treatments, as well as the exposure concentrations (e.g., more reference soil will be needed if many of the treatments are < 50% contaminated soil).
- If the study objectives require that screening tests be conducted, sufficient soil needs to be

Table 8. Minimum volume of bulk soil required for biological testing of a single soil sample

Test	Volume for one replicate (L)[a]
Nematode acute test	0.125
Oribatid mite test	0.5
Predatory mite test	0.6
Earthworm acute (Environment Canada method)	2.0
Earthworm avoidance (Environment Canada method)	5.0
Earthworm reproduction (Environment Canada method)	4.0
Enchytraeid reproduction	0.15
Collembola reproduction (Environment Canada method)	0.8
Snail test	0.25
Plant acute	1.5
Phytotoxkit	0.20
Plant definitive (Environment Canada method)	3.0
Plant life-cycle	2.4
Microbial respiration	0.05
Soil nitrification	0.05
Bait lamina	0.25
Community level physiological profile (CLPP)	0.05
Denaturing gradient gel electrophoresis (DGGE)	0.25
Soil enzyme assays	0.05
Physical/Chemical Analyses	
Particle size distribution	0.25
Water-holding capacity	0.1
Routine chemical properties (e.g., TOC, CEC, pH, SAR, EC, etc.)	0.5
Organic contaminants	0.25
Inorganic contaminants (e.g., metals)	0.25

[a]Note that these are minimum generic requirements for the tests listed and the ecotoxicity testing and chemical laboratories should be consulted for project-specific *sample volume* requirements. Extra soil should also be collected for contingencies as per guidance in Subsection 3.3.9.

collected for both screening and definitive testing.

- The amount of soil to be collected will depend on the soil volume/mass per replicate and the number of replicate requirements of the standard method(s) used.

- If the test replicates comprise soil *core samples*, rather than bulk soil samples, then the soil requirements can be very different.

b) The physical characteristics of the soil:

- The *bulk density* (mass of soil/unit of volume, usually g/cm^3) of the soil can significantly influence the amount of soil required for testing. The volume of soil in a test unit is the critical requirement for single-species soil toxicity testing, rather than the mass, since the soil volume defines the loading density of the organisms in a test unit.

- Soils with a high bulk density (e.g., sandy soils or heavy clay subsurface soils) might require a greater mass of sample compared with soils which have a low bulk density (e.g., peat or the organic layer of forest soils). For example, a single earthworm test unit that requires 270 mL of soil (for a 500 mL test vessel) could need up to 350 g (wet wt.) of sandy soil compared to only 200 g (wet wt) of peat.

- The moisture content of the site soil at the time of collection can also influence the amount of soil required for testing. Usually soil mass requirements in a test method are recommended based on the dry weight of the soil; therefore, if a site soil is very moist, more soil should be collected than if the soil at a site is dry.

- If the site soil contains significant amounts of large stones or industrial debris (approximately > 6 cm diameter), then more soil should be collected as these objects can comprise a significant portion of the collected sample, but will usually be sieved or manually removed from the soil samples prior to testing.

- If the site soil contains significant amounts of thatch or plant roots, then more soil should be collected as this material can also comprise a significant portion of the sample collected, and the amount of soil will be less once it is removed.

c) The distribution of the contaminant:

- For example, if the contamination is due to atmospheric deposition, the contaminant concentrations will be the greatest in the top few centimetres of the soil. If a few large samples are taken at depth (e.g., 0 to 30 cm) to meet the soil volume requirements for testing, after homogenization the contaminant concentrations in the test samples will be diluted and probably no longer represent the concentrations at the site. A better approach would be to collect many smaller samples at depths that represent the depth of contamination (e.g., 0 to 5 cm) and homogenize these samples to meet the soil volume requirements for testing (see also Subsection 3.3.10.1).

3.3.10 Sample Number

The number of samples to collect depends upon the study objectives, the data quality objectives, the desired level of certainty, and site-specific considerations such as the distribution of the contaminants, the heterogeneity of the soil, the soil testing requirements, and the size and homogeneity of the study site. The number of samples collected at a site is usually a compromise between the requirements of the data quality objectives and practical constraints such as sampling and testing costs (Mason, 1992; EC, 1994; Bélanger and Van Rees, 2008).

In order to calculate the number of samples that need to be collected for a study the following questions must be answered:

- The level of certainty required when making a management decision based on a statistical sample. Whenever a statistical sample is used to make a decision there is always a chance of making one of two types of errors. A type I error occurs when a null hypothesis is incorrectly rejected. For example, if the null hypothesis is the absence of an environmental "effect" then the conclusion "There is an environmental effect" will be incorrect. A type II error occurs if the null hypothesis is accepted when it should have been rejected. Again, for example, if the null hypothesis is the absence of an environmental "effect" then the conclusion "There is no environmental effect" would be incorrect. Accepted practice when making environmental decisions is

Canada is to set the two error rates equal to one another and usually less than 10%. Note that a specific statistical power is often prescribed as a criterion for adequacy of a site assessment. Statistical power is equal to 1 – Type II error. A reasonable number for the statistical power is preferably 90% or higher and certainly no less than 80%.

- Precision around some measurement. At times there is interest in measuring the degree of "effect" in a soil with a pre-specified level of precision.[21]

Once the level of certainty required for a site assessment has been decided sample size calculations can be performed. These calculations require some estimate of the variability in the response; this might be obtained using preliminary/historic data, best professional judgment or a worst-case scenario.

Usually, the estimated sample size will be a compromise between what is desirable and what can be achieved. If a very high level of certainty is required and the response is variable a very large number of samples might be required. Conversely, if a lesser degree of certainty is required and variability is low or can be adjusted by a judiciously chosen sampling strategy, a lesser number of samples may be required.

Finally it is important to consider the nature of change that constitutes a deleterious effect when estimating sample sizes. For example, some might consider *any* significant change in a response at a site relative to a control deleterious, whereas others would argue that say, an increase in root length at an exposure site is not a deleterious effect. If a change in only one direction is considered deleterious then sample size calculations should be based upon one-sided hypothesis tests. Note the decision regarding what comprises a deleterious effect should be made prior to interpreting data.

The answers to these questions regarding certainty, precision, and what constitutes a deleterious effect must be specified in the data quality objectives (Subsection 3.3.1). One of the most common scenarios for biological testing is to determine if the mean value for a contaminated site differs significantly from the mean for a reference site. To determine the number of samples required to achieve the precision and accuracy specified in the DQOs for the comparison between two means, where the data used to estimate the means follows a normal or Gaussian statistical distribution, the following generic equation (for a two-tailed test) can be used (Mason, 1992):

$$n \geq 2\,[(Z\alpha + Z\beta) / D]^2 + 0.25\,Z^2\alpha \qquad [1]$$

where:

n = number of samples

$Z\alpha$ = Z statistic for Type I error probability (e.g., $\alpha = 0.05$)

$Z\beta$ = Z statistic[22] for Type II error probability (e.g., $\beta = 0.90$)

D = minimum relative detectable difference/CV

CV = coefficient of variation

However, this is a generic example only. The calculations used to determine the number of sample can vary depending on the sampling strategy used, the statistical distribution of the observed data, the presence of spatial correlation and the questions being asked (e.g., comparison among multiple sites versus comparison between two sites). Many of the available equations require preliminary site data; however, when conducting biological testing for a contaminated site, these data are usually available. Appendix C provides more detailed guidance for calculating the number of samples using examples from simplified case studies that

[21] In some cases, this is formally defined in a protocol or regulation. For example, one of the criteria for the site-specific ecotoxicity testing Pass/Fail approach in the Alberta Environment draft Tier 2 Eco-contact Guideline Derivation Protocol (AE, 2007b) is that the experimental design must have adequate power to detect a difference of 25% or more between treatments.

[22] The Z tables needed to do the calculations in Equation 1 are provided in Appendix J.

represent different potential study objectives and sampling strategies.

It is usually worthwhile and cost-effective to consult a statistician when determining the sampling strategy to use and the number of samples to collect at the outset of a study.

3.3.10.1 Point, Composite, and Bulk Samples

Point samples (also called sample increments), are individual blocks of soil removed from one sample location by a sampling device (e.g., a soil core). *Composite samples* are samples comprising of two or more point samples. When point samples from different sampling locations are pooled together, the pooled sample is a composite sample. Bulk samples are large (e.g., > 1 L) point samples that consist of more than one individual blocks of soil removed from one sample location by a sampling device; often collected to satisfy the large volume requirements for biological testing. Bulk samples are often collected from sample locations that are 1 m^2 or more, and are usually collected by horizon or depth. Although soil might be removed from the discrete (e.g., 1-m^2 plot) location more than once (e.g., using a *corer*), it is the entire volume removed from that designated sample location (e.g., the entire *A horizon* of a 1-m^2 plot) that is considered the point sample (EC and SRC, 2007). A number of studies have shown that the results of bulking samples from individual sample locations can provide good estimates of the real soil contaminant concentrations and soil properties in forest soils; soils that are characteristically very heterogeneous (Bélanger and Van Rees, 2008). When one or more bulk samples are combined, (e.g., bulk samples from two or more sampling locations at a site are combined), the result is a composite sample with a very large soil volume.

Collecting composite samples is one way to decrease the variability in contaminant concentrations and soil characteristics in soils tested (USEPA, 1993). Composite sampling is generally appropriate at sites in which there is a homogeneous distribution of contaminants and soil properties, or, more commonly, the distribution is heterogeneous over a small-scale, but at a larger-scale the distribution is considered to be more homogeneous (ISO, 2005b). Point samples from the entire site may be composited, or separate composite samples can be made from different areas of the site (Athey *et al.*, 1987). Compositing samples can produce more representative samples of the site, and often provide an excellent estimate of the mean of the pooled point samples.

The primary disadvantage of collecting composite samples is that they cannot be used to: estimate the variability of the toxicity; detect the toxicity of the highest soil contaminant concentration; or, determine the influence of the variability of soil characteristics on toxicity (ISO, 2002a). Because diluting hot spots is always a possibility when collecting composite samples, testing composite samples does not necessarily result in testing soil samples representative of the "worst-case" or "hot-spot" scenario, that is, soils that are most contaminated and assumed, therefore, to be most toxic. The loss of information about the variability of the soil physical and chemical characteristics is also important because these can strongly influence soil toxicity. Site assessors and managers should be aware that collecting composite samples can result in a less conservative approach and is often undesirable from a regulatory perspective; therefore, it should be discussed with regulators in advance of soil collection. Some composite samples can also be difficult to homogenize (e.g., if *subsamples* are moist and clayey) (Hazardous Waste Consultant, 1992; CCME, 1993a; Mason, 1992). Collecting composite samples is also not appropriate when soil samples must remain undisturbed, such as when soils are contaminated with volatile compounds or are being collected for microcosm studies with intact cores (ISO, 2002a; USEPA, 1993).

One approach that can be used to overcome some of the limitations of composite sampling is if the toxicity of a composite sample exceeds a pre-specified criterion (e.g., > 25% difference from the reference soil) (adjusted for the dilution of the point samples by compositing), then the point samples that made up the composite samples could be tested again separately (Athey *et al.*, 1987). This assumes of course that subsamples of the point samples used to make up a composite sample were stored before

compositing. This approach would require large volumes of soil to be collected for a single composite sample. A method to estimate the maximum number of point samples *(n)* that should be grouped into a composite is proposed by Athey *et al.* (1987) as:

$$n \leq MAL/MDL \qquad [2]$$

where:

n = maximum number of point samples

MDL = the minimum detection limit for the biological tests (e.g., a measure of the variability of the control samples)

MAL = the maximum acceptable limit for the contaminant (e.g., the maximum percent difference from the control that is considered not to represent an adverse effect)

Upon arrival at the toxicity testing laboratory, unconsolidated soil samples (point, bulk, or composite) are usually homogenized and subsamples are tested as a soil treatment in a toxicity test (e.g., all the test replicates originate from one field sample). If only one field (point, bulk, or composite) sample is tested per site, no information is provided about the variability of the toxicity at the site (as testing subsamples provides information about the variability of the test organism response to soil from one sample only) and statistical comparisons of soil toxicity among more than one site cannot be made. If more than one field sample is tested per site (regardless if the field sample is a point, bulk, or composite sample), then the variability of the toxicity of the site can be estimated and comparisons of soil toxicity among sites can be made.

Rather than collecting large bulk point samples and testing subsamples in a laboratory test, small point soil samples that consist of only one individual block of soil removed from a sample location can also be collected for biological testing and can provide a cost-effective alternative to collecting and testing multiple bulk samples. Point samples can be unconsolidated (e.g., collected with a large soil corer, removed and placed into a sample bag or bucket in which the unconsolidated soil is shipped to the laboratory) or consolidated (e.g., collected with a large soil corer with a liner, removed from the corer intact, liner sealed, and then shipped to the laboratory intact). Each small point soil sample can then be tested as a single laboratory test replicate if the volume of soil is adequate for the testing required. Because these laboratory test replicates represent true field sample replicates and provide information about the variability of the toxicity at the site, this approach can considerably reduce the volume of soil to be collected and tested (i.e., multiple bulk soil samples do not need to be collected) and the toxicity of multiple sites can still be statistically compared. The disadvantage of this approach is that the inherent variability of the biological response of test organisms can be high (the use of subsamples minimizes this variability) with the result that the variability of test organism response among multiple point samples within a site might prevent the detection of differences in the response of biological test organisms between sites. Sampling and laboratory experimental designs that employ combinations of testing point samples (field sample replicates) and subsamples of bulk or point samples in a toxicity test can be used to address these different concerns.

3.3.11 Quality Assurance and Quality Control

Quality assurance (QA) is a system of management and operational activities designed to ensure adequate control of the quality of the work being performed. Quality control (QC) is part of QA and involves routine checks and calibrations in normal operations (EC, 1999). The goal of QA/QC programs is to identify, measure, and control the errors associated with every component of a sampling study, including planning, sampling, testing, and reporting. Detailed QA/QC program and project plans should be a part of any soil sampling study. Quality assurance effort should be focused on the following activities:

- setting the DQOs
- developing a sampling strategy that meets the DQOs
- developing a sampling plan that meets the DQOs, including collecting sufficient soil samples and replicate samples

- ensuring that the reference soil selected is appropriate and the physical and chemical characteristics match the contaminated soil as much as possible
- ensuring that qualified personnel (e.g., soil scientists or other experienced personnel) collect soil samples
- ensuring the use of standardized field sampling forms
- ensuring that acceptable sample handling, preparation, and manipulation procedures are clearly outlined in the study plan
- ensuring that appropriate biological test methods are used
- ensuring that ecotoxicity and analytical laboratories are experienced and accredited; it is recommended that laboratories hired are accredited for the tests they will conduct by the Canadian Association for Laboratory Accreditation (CALA) [formerly Canadian Association of Environmental Analytical Laboratories (CAEAL)], Standards Council of Canada (SCC) or by the Centre d'expertise en analyse environnementale Québec (CEAEQ) for Quebec-based laboratories. The use of CALA-, SCC- and/or CEAEQ- accredited laboratories will ensure that appropriate standard operating procedures are followed, that documented QA/QC procedures are in place, and that the testing is conducted by well-trained and experienced personnel.

Quality control activities include:

- documenting that the sampling plan was followed
- documenting when deviations from the sampling plan occur and providing rationale for the deviations
- taking detailed field notes and observations
- properly collecting and handling samples through use of proper sampling devices and standard operating procedures
- properly decontaminating soil sampling equipment between sampling locations and sites
- properly decontaminating field personnel and auxiliary equipment (e.g., boots or containers) between sampling locations and sites
- proper transport and receipt of soil samples at the laboratory
- proper documentation of custody and transport of sample shipment to the laboratory
- storing, preparing, and manipulating samples properly in the laboratory
- taking detailed notes and observations of sample storage, preparation and manipulation in the laboratory
- documenting when deviations from the study plan occur regarding sample storage, preparation, manipulation, and testing in the laboratory and providing rationale for the deviations

3.3.11.1 Preventing Sample Cross-contamination

Sampling equipment should not be made of material that will contaminate the sample (e.g., sampling devices should not be painted, plated, greased, or have some other type of chemically treated surface) (Table G.1). Stainless steel is usually the material of choice, and inert plastic equipment can also suffice in some cases when collecting metal-contaminated soils (ISO, 2002b). Prevention of cross-contamination in the field is essential, and the level of effort required will depend upon the characteristics of the contaminant(s) of concern. Sampling devices should be cleaned between sampling locations and sites and at the end of sampling by:

- thoroughly wiping, washing, or scrubbing the equipment;
- rinsing the equipment with purified water brought to the sampling site;
- equipment can also be cleaned with non-phosphate containing soap, but the equipment must be thoroughly rinsed following a soap wash to ensure that the samples do not become contaminated with soap residue;
- if the samples are to be used to collect soil containing organic contaminants, rinsing the equipment 3× with a solvent (acetone or

hexane) of appropriate grade (then discarding contaminated solvent into a waste container for later disposal according to applicable environmental and waste management regulations)[23] and then air-drying the equipment;

- rinsing equipment twice more with purified water; and,
- package equipment in plastic bags (or aluminum foil if there is potential that organic contaminants could leach from the plastic bags and cross-contaminate the sampling equipment).

Not all of these decontamination procedures may be required for each sampling scenario; the specific procedures to use should be decided on a project-specific basis. Individual sets of sampling equipment can be used for different sampling locations and/or sites to reduce the number of times sampling equipment needs to be cleaned between sampling events (ASTM, 2008a). Personal protective equipment (e.g., boots) should also be cleaned in a similar manner, and care should be taken to not step on the soil surface within the sampling location if possible. Care should also be taken to prevent cross-contamination of equipment and samples from auxiliary substances such as glue, grease, fuels, exhaust fumes, etc. (ISO, 2002b).

3.3.11.2 Sample Replicates, Split and Blind Samples

The number of sample replicates to collect in any study is dependent upon the study objectives, experimental design of the biological tests, logistical and budgetary constraints of sampling, and the cost of sample collection, preparation, manipulation, and testing. It is recommended that more than one point, bulk or composite sample be collected per site to provide information about the variability of the toxicity/bioavailability of the contaminants at the site; and so that statistical comparisons of soil toxicity can be made among more than one site. Contingency samples (e.g., more samples than actually needed) can also be collected if time and logistics allow.

Split samples are created when one point, bulk, or composite sample is split into two equal parts immediately after sampling (Figure G.5; Appendix G). Each split sample is handled and prepared in an identical fashion and then submitted to the testing laboratory as individual samples. At the laboratory each sample subjected to the same analysis (e.g., biological testing) and the results are compared. Split samples are not replicate samples; the only variability expected in the results between split samples is that of acceptable laboratory test variability. If the variability in the results between the split samples is greater than expected this indicates that the sample collected was not representative, either due to the physical techniques used to collect the sample or due to the variability of the soil at the sample location (or, less likely, due to unacceptable laboratory test variability. Examination of laboratory QA/QC data and discussion with laboratory personnel can clarify this potential issue).

Blind samples are samples that are submitted to a testing laboratory without any identifying information that would link the sample to a specific location, physical-chemical characteristics, or nature or degree of soil contamination. Samples submitted without identifiers ("blind" samples) are usually replicate or split samples. The purpose of submitting blind samples is to confirm that the results generated by the testing laboratory are not influenced by prior knowledge of the sample. This is a common practice for chemical characterization of a site but is less commonly used when samples are collected for biological testing. This is partly due to the fact that the volume of soil and cost of testing individual samples can be significantly greater for biological testing and because it can significantly inhibit proper interpretation of the biological test results. If blind samples are to be submitted as part of a study, early discussion with the testing laboratory about the possible receipt of blind samples is recommended.

Other environmental QA/QC samples such as trip blanks, equipment blanks, field spikes, and

[23] This method is rarely used because of the safety hazard that solvents such as acetone and hexane represent. If organic solvents must be used, personnel handling these solvents should wear respiratory protection.

reference materials, widely used in studies for chemical characterization of contaminated sites, are not applicable for collecting soil samples for biological testing and are not described in this document.

3.3.12 Environment, Health, and Safety

The health and safety of all personnel sampling and handling contaminated soil must be ensured. Hazards on-site include physical hazards such as unstable ground, slopes, open holes, or excavations. Injury from mechanized equipment, mishandled sampling devices or strain from overexertion can also occur (ISO, 2002b). Exposure to contaminants from inhalation of soil dust or vapour, through dermal contact, or through ingestion of soil particles can also cause acute or chronic injury to sampling personnel or people within the vicinity (ISO, 2002b). In some situations risk through exposure to biological hazards may be present. Specific hazards exist when sampling soils at a site contaminated with residues from explosive, chemical, or biological warfare, or radioactive materials.

All possible hazards must be considered in the study plan when selecting sampling methods, devices, and the most appropriate personal protective equipment for field personnel. A safety policy must be in place before personnel travel to the field, and adherence to the policy and its standard operating procedures is mandatory (ISO, 2001). Smoking, eating, or drinking during sampling, handling, or testing soil samples must not be permitted (ISO, 2001). Appropriate safety footwear should be worn, and protection taken against physical risks including alertness and caution when traversing contaminated sites. Machinery should only be operated by qualified personnel; risks from wildlife can be mitigated by carefully planning the time of year to sample, carrying deterrents (e.g., bear spray), and working in a group. Sampling should never be conducted alone. Personal protective equipment (PPE) suitable for protection against chemical and biological risks includes chemical-resistant disposable gloves and chemical-resistant safety boots and coveralls (impervious when appropriate). If chemical hazards exist on-site, then safety glasses, goggles, or full-face protection might also be necessary. If toxic gas is expected to be present on-site, then portable gas monitors should be issued to personnel to ensure the absence of hazardous gas concentrations (ISO, 2001). In some cases, full protective clothing including an external air supply might be required. In addition to the above-mentioned PPE, to protect against the risk caused by exposure to bacteria and viruses, tetanus, hepatitis, and typhoid vaccinations should be kept up-to-date for all sampling personnel. At sites where soils are contaminated with residues from explosive, chemical, or biological warfare, or radioactive materials, safety training and measures (including PPE) should be provided by experts, and sampling should be conducted in the presence of these experts.

Laboratory personnel handling, preparing, and testing the soil should always be informed of the potential contaminants of concern and their estimated maximum levels present in the soil, or suspected of being present in the soil, *before* their receipt at the laboratory. Laboratory personnel can then take appropriate measures to minimize their risk through control measures, the use of appropriate equipment and handling procedures, and the use of personal protective equipment.

Sampling activities should be carried out carefully so that contaminated dust does not become stirred up with the potential to move to other parts of the site (ISO, 2002b). When sampling equipment is decontaminated, all rinseate (water and solvent) should be collected in waste containers and disposed properly off-site following sampling activities.

All garbage and debris resulting from sampling must also be collected and disposed of properly off-site and machinery and wheels of vehicles should be cleaned prior to leaving a site to prevent the spread of contamination (ISO, 2001). When sample testing is finished at the laboratory, waste soil and biological material must be disposed of properly according to local environmental and waste management regulations.

3.3.12.1 Backfilling Excavations

In some cases, soil sampling for biological testing can result in a substantial volume of soil removed from a sampling location, for example, when test pits are created. Once the soil has been removed, the excavated area must be properly backfilled to prevent it from being a physical hazard on the site, a preferential pathway for contaminant movement within the soil, or to return the sampled area to pre-sampling use (e.g., agricultural or parkland use). Excavated soil and other material should be segregated (e.g., into topsoil, subsoil, and parent material) and replaced in the order in which they were removed. Not only must the material be returned in the same order, the bulk density should be comparable to unexcavated material; therefore, the material might require some compaction after backfilling (keeping in mind that materials should be left loose within the root planting zone as compaction might affect plant growth). If frozen soils are used for backfilling, action should be taken to prevent future subsidence. If substantial quantities of soil are collected, then soil might need to be imported into the site for backfill. The physicochemical characteristics of the backfill material should be comparable to the original material; for example, the soil texture, pH, salinity, sodicity, *total organic carbon* content, should be similar, especially if materials are imported from another geographic location. The quality of the backfill might also be of concern as some jurisdictions might require prior landowner approval for the importation of backfill material onto their property. Quality of the backfill might not be restricted to chemical constituents, for example soil might need to be demonstrated to be free of biological contaminants in regions where such agents are of agricultural concern (e.g., clubroot in Alberta). The physical, chemical and, when required, biological, characteristics of the backfill material should be determined by laboratory analysis prior to use on the site.

3.4 Sampling Plan and Preparations for Field Sampling

A sampling plan is a key component of the study plan. A sampling plan is a written description of the detailed procedures to follow when collecting samples, handling and preparing samples on-site (if required), packaging, labelling, storing (if necessary) and transporting samples. The sampling plan includes the documentation associated with each activity. A sampling plan is an important supporting document for a study plan and should include all of the information listed in a Sampling Plan Checklist (adapted from CCME, 1993a) (Table 9).

3.5 Field Measurements and Observations

The sampling plan should include explicit instructions on the field measurements and observations to be documented. Generic recommendations for observations and measurements are provided in the Field Notes Checklist adapted from CCME (1993a) and ISO (2002a) (Table 10); however, it is expected that these requirements will be modified to meet site-specific needs.

3.6 Collection of Soil Samples

There are two general kinds of soil samples: consolidated and unconsolidated. The most commonly collected type of sample for biological testing is unconsolidated; with these types of samples soil particles become loosened and separated in the sampling process. Consolidated samples are those collected such that the soil particles and pore structure remain unaltered compared with the original ground structure (ISO, 2002b). Consolidated samples are collected for biological testing when soils are contaminated with volatile compounds, for microbial testing with consolidated cores, for single-species toxicity testing with consolidated cores and, for microcosm testing with consolidated cores.

3.6.1 Soil Classification

Prior to extracting soil samples, it is important to obtain a thorough field description of the soil to be sampled, as this provides a basis for the proper collection of samples as well as the

Table 9. Sampling plan checklist

- Site location(s) (and directions to get there)
- Site manager contact information
- Proposed sampling locations

- DQOs and procedures required to meet them
- Modifying protocols in case problems are encountered in the field (including changing site locations or minimum sample volumes required, sampling equipment failure, sampling personnel substitutions, or the occurrence of extreme weather)

- Identifying sample locations (e.g., installing stakes)
- Preparing the site (e.g., marking plot boundaries, removing litter layer)

- Sampling devices (e.g., types, number)
- Samples [including sample volume, depth(s), horizon(s), number of bulk, point, or composite samples, number of subsamples for composite samples, number of replicates]
- Sampling procedures (e.g., compositing, special times or conditions to sample, contaminant-specific sampling procedures)
- Sampling containers (including types, material, size, number, labels)

- Other sampling material including equipment for field observations, field measurements and sample preparations in the field (e.g., field logs, GPS unit, camera, labels, coolers, icepacks, tape, waterproof pens, packaging material, sieves, tarpaulin, rakes, stakes, spray paint, waste disposal containers)

- Field preparations of sampled soil (e.g., sieving, homogenization, drying)
- Field measurements to make
- Field observations to record

- Sample packaging, transporting, and storing
- Documentation forms and procedures (e.g., sample labels, access permits, *chain-of-custody*, transportation of dangerous goods, sample seals)

- Decontamination/hygiene procedures (e.g., equipment, personnel, *sample containers*)
- Waste disposal procedures

- Environment, health, and safety equipment (e.g., PPE, appropriate clothing, sunscreen, insect/wildlife repellent, decontamination water and/or solvent, brushes, cloths, waste containers)

- Safety plan (e.g., emergency procedures, emergency contact numbers, communication equipment)

Table 10. Field notes checklist

Soil sampling
- sampling date
- sampling time
- sample identification number
- sample location
- sampling site
- sampler's name
- sampling conditions
- sample type
- sampling device
- sample volume

Sample handling
- sample preparation (e.g., sieving, drying, homogenization)
- *subsampling* for chemical analyses
- sample observations (e.g., odour, indigenous fauna, moisture status)

Field measurements (ancillary measurements to be taken)
- co-located sample collection for contaminant or soil physicochemical analyses
- *in-situ* testing of soil properties (e.g., bulk density pH, soil moisture)
- soil horizon characterization (if not previously characterized)
- *in-situ* field testing (e.g., litterbag, earthworm exposure, bait lamina, DNA probes)
- ecological survey(s)
- field sampling of flora and fauna

Sample site observations
- weather conditions (e.g., air temperature, rainfall, sunny, cloudy)
- visible contamination or areas with no vegetation
- photographs of site, soil profile and sample locations
- description of soil profile or layers
- sketches of site and sample locations
- infrastructure near sample location
- topography and hydrological characteristics (e.g., slope, presence of surface water)
- identification of dominant vegetation (e.g., trees, shrubs and herbs) at site (if not done previously)

Sample storage
- storage date, time, and conditions
- storage duration

Sample transportation
- shipment date
- shipment time
- name of sampler relinquishing sample(s)
- transport method and service supplier
- tracking number for transport
- transport condition
- chain-of-custody forms
- transportation of dangerous goods forms (if required)
- confirmation and date of receipt at laboratory

interpretation of biological test results. Soil surveys and maps provide information about the soil at a regional level, as described in Subsections 3.3.3.2 and 3.3.3.3. However, soils collected for biological testing should also be described at a detailed site-specific level. Soils are identified and classified within a hierarchical taxonomic system in much the same way as plants and animals. The concept of soil properties (chemical and physical parameters) reflecting the action and interaction of soil forming processes over time allows soils to be classified and related based upon their soil properties, and not just the environmental factors that influence the soil (e.g., vegetation).

In Canada, soils are classified using the Canadian System of Soil Classification (CSSC) (AAFC, 1998)[24] based upon the similarities and/or differences of their soil properties, which reflect the action and interaction of processes over time. Soil-forming processes are factors that change and/or determine the rate, direction, and extent of change of the parent rock material to form soil. There are five recognized soil-forming factors: parent material, climate, living matter, topography, and, time (Hausenbuiller, 1985). These factors form soil by continuous mineral weathering and clay synthesis, organic matter accumulation, exchange of ions, translocation of soluble and soil components with the profile, structure formation and mixing of soil materials (Hausenbuiller, 1985). The cumulative effect of these processes also results in the development or degradation of soil horizons. Soil horizons are layers of mineral or organic soil material approximately parallel to the land surface that have characteristics altered by processes of soil formation (AAFC, 1998). Soil horizons are typically observed within soil profiles as lateral layers of mineral or organic material. The primary mineral horizons are defined as A, B, and C; the primary organic horizons are L, F, and H (predominantly forest litter), or O (predominantly wetland vegetation). A soil profile is typically what is observed and/or sampled in the field to determine the classification of a soil. An illustration (Figure E.3) and description of a soil profile is provided in Appendix E.

Soils are classified using the CSSC in order of increasing specificity and taxonomic relation, as Order > Great Group > Subgroup > Family > Series. The Soil Order reflects both the effects of the dominant soil forming processes and/or environmental factors. There are 10 main Soil Orders within the Canadian landscape, most of which predominantly occur in one of the following major ecological systems: forest, grassland, or tundra, and as such, have a defined geographic extent. A detailed description of each these Soil Orders is provided in Appendix E, which includes diagnostic horizons, photographs, and schematics of their geographical extent in Canada.

Soils collected for biological testing should be classified by a soil scientist or other experienced personnel at a minimum to the Subgroup level but preferably to the soil Series level. A Series contains soils that "have similar kinds and arrangements of horizons, whose colour, texture, structure consistence, thickness, reaction and composition fall within a narrow range" (AAFC, 1998). Soils within a Series are identified by a specific name, which is typically linked to a geographic area; the name becomes representative of all of the characteristics of a particular soil. At the field level, a soil Series is associated with the soil profile, its classification, and the location of that profile within the topographic landscape. Readers should refer to Appendix E for more information on the CSSC and the basic components of soil taxonomic identification.

[24] Other methods of soil classification are used in Canada for purposes other than soil science, agronomy, or biological testing. The Unified Soil Classification System (ASTM, 2006a), the system described in the Canadian Foundation Engineering Manual (CFEM, 2007), and the geological assessment method described in Compton (1985) are commonly used by engineers and geoscientists for contaminated site assessments. Since biological testing might be conducted with soil that has been characterized using these other classification systems, it is important to confirm that soil samples collected for testing have also been classified according to the Canadian System of Soil Classification.

3.6.2 Obtaining Reference Soils for Biological Testing

One of the most important tasks, and sometimes the most challenging, when conducting site-specific biological testing is the collection of a well-matched, site-specific, field-collected reference soil(s). A site-specific field-collected reference soil is a one that has physical and chemical properties similar to those of the contaminated field-collected soil(s), but is not contaminated.

The importance of including a reference soil as a negative control soil in a site-specific ecotoxicity assessment is that it enables the investigator to differentiate between the influence of the soil's physical and chemical properties on test organism performance from that of soil contaminant(s). In many cases inherent soil properties, such as texture, pH, electrical conductivity, fertility, and organic matter content strongly influence the growth and vigour of plants, the growth and reproduction of invertebrates, and the activity and diversity of soil microorganisms. In order to avoid a misinterpretation of the toxicity data, therefore, well-matched reference soils should be included in any site-specific ecotoxicity assessment. Most types of site assessments (including microcosms) require the inclusion of reference soil treatments; however, the ecotoxicity laboratory conducting the tests should be consulted during the planning stages regarding the nature and amount of reference soil needed to meet the study objectives and DQOs.

Selecting a well-matched reference soil can entail significant effort and should be given due control site or area to collect reference samples should be conducted. Reference soils should be classified as described in Subsection 3.6.1. In addition to being from the same soil Series as contaminated soil, reference soils should also be from the same soil horizon and have undergone the same level of disturbance (e.g., soil compaction, clear-cut forest site, etc.) as contaminated samples (EC, 1999). Some of the most critical soil physicochemical properties that should be matched between the reference and contaminated soils include:

- particle size distribution (percentages of clay, sand, and silt)
- organic matter content[25]
- pH
- electrical conductivity
- fertility
  - nitrogen as total N, nitrate (NO_3^-), nitrite (NO_2^-) and ammonium (NH_4^+)
  - potassium as plant-available or total potassium
  - phosphorus as plant-available and total phosphorus
- C:N ratio (for microbial tests)

In addition, other properties to match might include:

- *cation exchange capacity*
- total carbon content
- total inorganic carbon
- exchangeable cations (Ca^{2+}, Mg^+, Na^+, K^+)
- oxidation-reduction (redox) potential
- water-holding capacity (especially for microbial tests)

Other soil properties may be critical to match depending on the type of biological testing to be conducted. More than one reference soil might be required to match the contaminated soil samples, especially if there is a wide range of physicochemical properties among site soils.

Determining if a reference soil is well-matched to the contaminated soil must be conducted on a case-by-case basis. General guidance on the degree to which a difference in soil physicochemical properties between contaminated and reference soils is considered acceptable is difficult to provide as it will vary with the type of soil, toxicity test(s) being performed, and test organism(s) chosen. In addition, the acceptability of differences between physicochemical characteristics of reference and contaminated soil will vary among individual soil characteristics (e.g., acceptable

[25] Organic matter content is calculated from total organic carbon (TOC) and can be estimated from TOC by multiplying the organic carbon content of a soil by 1.72 (Hausenbuiller, 1985). However, the relationship between TOC and OM is slightly different among soils and the organic matter content and, therefore, should also be determined by laboratory analysis.

difference in % silt is probably greater than acceptable difference in % organic matter). Test species also will have different ranges of tolerance for changes in soil characteristics [e.g., some invertebrate species might be sensitive to changes in pH, whereas some plant species may have a wide range of pH tolerance (Jänsch *et al.* 2005)]. Soil characteristics influence both the test species performance and contaminant bioavailability. Matching soil physicochemical characteristics of the reference soil to those of the contaminated soil is a compromise between matching some key physicochemical characteristics at the expense of other physicochemical characteristics that are not considered to have as strong an influence on contaminant bioavailability and/or test species performance.

In addition to matching key physical and/or chemical properties of the reference soils to those of the contaminated soils, the investigator needs to ensure that the reference soils themselves are not contaminated. To screen candidate reference soils for possible contamination, the following minimum chemical and biological analyses are recommended:

- organophosphorus insecticide suite
- organochlorine insecticide suite
- herbicide suite
- metals suite
- petroleum hydrocarbons (including PAHs)
- other site- or area-specific contaminants of concern

Performance tests should be conducted with the test species to be used for the study to ensure the reference soil adequately supports survival, growth, and/or reproduction of the test species (EC, 2004a, 2005a, 2007a). The chemical and biological methods used should be in accordance with the requirements of the jurisdiction under which the study is being conducted.

Reference soils should be collected adjacent to, or in the vicinity of, the study site. If this is not possible, then reference soils should be located in a region similar to that of the study site. Regardless, the reference sites should be located upwind (for aerial contamination) or up-gradient (for groundwater-mediated soil contamination) of the sampling site. Travel between reference and sampling areas should be minimized and the reference site samples should be collected before the contaminated site samples when possible and appropriate to do so (e.g., weather conditions or access to a site might dictate otherwise). Reference and contaminated soil should be collected from areas with similar topography where possible. The use of local reference sites is preferred; however, when an appropriate local reference site is not available, an area reference site is acceptable. Sample collection and QA/QC techniques at reference sites must be identical to those at sampling sites.

Sometimes site-specific reference soils cannot be located for a study; this is why it is critical to try to locate a reference soil early in the study process so DQOs and study objectives can be modified, if necessary. If matched reference soils cannot be located for a study, a generic reference soil that matches the site soil(s) as best as possible can be purchased (ISO, 2003b, Römbke and Amorim, 2004)[26]. Currently there are no commercial sources of field-collected soil available in Canada for biological testing; however, for information on future availability of such soils it is recommended that a soil ecotoxicity testing laboratory or the Environment Canada Science and Technology Directorate personnel referenced at the front of this guidance document be contacted. Alternatively, when a suitable reference soil

[26] In Europe, ecotoxicologists have the option to test with a series of "certified" natural soils collected from specific regions. These soils are extensively characterized and provide investigators with a natural soil that can be used, and results compared, among different studies. These soils are uncontaminated and can, therefore, be used as a generic reference soil. The most commonly used standardized natural reference soil used in Europe are the LUFA (Landwirtschaftliche Untersuchungs und Forschungs Anstalt) soils collected in Germany obtained from Landwirtschaftliche Untersuchungs und Forschungs-Anstalt Speyer, Obere Langgasse 40, D-67346 Speyer, Germany. With any soil bank, this source of certified natural soil will become depleted over time; therefore, it is still best recommended that every effort be made to collect a local or area field reference soil.

cannot be located for test designs; incorporating samples with concentration gradients of contaminants, the use of multivariate techniques (Jensen *et al.*, 2006b; Strandberg *et al.*, 2006), or spatially explicit techniques that use geostatistical tools to determine spatial correlations among concentrations of contaminants and biological endpoints (Kuperman *et al.*, 1998; Subsection 3.3.6; Appendix D), can overcome this problem.

3.6.3 Collection by Soil Horizon

Most Canadian soils are highly stratified into soil horizons, which develop through deposition of material on the surface and through downward leaching of water-soluble constituents. As a result, the structure and chemistry of soil horizons are often very different and this can result in different bioavailability and toxicity of contaminants to soil organisms. Soil contamination can be stratified according to soil horizon, in part due to the different speciation and resultant mobility of contaminants in different horizons. Soil sampled for biological testing should in most cases be collected by soil horizon (ISO, 2002a, 2006a) and/or when there is a change in material type, colour, and/or texture. There are exceptions to this general recommendation, however, and they are provided in Subsection 3.6.4. Although agricultural soils that are cropped are disturbed by tillage, the resulting disturbed layer is identified as a horizon (the Ap horizon, which is the surface horizon disturbed by human activity such as tillage, logging, or habitation) in the Canadian System of Soil Classification and the Ap horizon should be sampled as its own soil unit. Sampling soils by horizons generally results in less variability because the mixture of soil material with different properties (and/or contamination) is avoided (Bélanger and Van Rees, 2008).

The top layer (A horizon) is the most commonly sampled horizon for biological testing. This horizon contains the most organic matter and most of the biological activity in *mineral soil*; in addition, aerial deposition and surface spills elevate contaminant concentrations in this horizon. Depending on the study objective, the forest litter (L layer), fulvic/humic (FH horizon) (e.g., at a forested site), or surficial organic layer (*O horizon*) of mineral soils (e.g., at a tundra site) might also be collected, when present. Subsurface *B horizons*, and less commonly, C, horizons are also sampled, when soil contamination is observed or measured in these layers. Contamination is present in these horizons either as a result of downward leaching of spills, upward leaching from contaminated groundwater, or direct contamination through anthropogenic activities.

To sample soil by horizon, the soil profile at the site must first be classified as described in Subsection 3.6.1. Soil profile characterization can be done either during earlier stages in the site assessment or at the same time as samples are collected for biological testing, as described in Subsections 3.3.3.2 and 3.6.1. Historical records and background soil data for the site provides useful preliminary information and if sufficiently detailed, may preclude the need to characterize the soil profile on-site (Subsection 3.3.3). If this background information is not available, tests pits could be dug to observe the soil profile. These pits vary in depth depending on the site but often range from 0.5 to 1 m. As an alternative to digging test pits, an estimation of the horizon depths can be obtained from soil cores withdrawn from within or adjacent to the sample location prior to sampling. For example, if a sample location is a square plot, soil cores could be removed from two opposite corners of the plot; or if the site location slopes or has other topographical features or the site is highly heterogeneous, four cores could be removed, one from each corner. Soil profile characterization should be carried out by a soil scientist or experienced personnel.

Care should also be taken when sampling according to soil horizon that dilution of the soil contamination does not occur. For example, in the case where it is known that the vertical contamination extends only partially through a soil horizon (e.g., contamination extends from the surface through the A horizon to a depth of 20 cm in the B horizon, but the B horizon ranges from 10 to 40 cm) then the contaminant concentrations in the B horizon will be diluted. In a situation such as this, a combination of

sampling by horizon and depth can be used [e.g., the A horizon (0 to 10 cm) and B horizon (10 to 40 cm) are sampled separately, but the B horizon is sampled only to a depth of 20 cm, or the B horizon is collected as two different samples at two different *sampling depths*: 10 to 20 cm and 20 to 40 cm].

3.6.4 Collection by Soil Depth

Collection of soil samples according to depth is recommended for soils without distinct soil horizons (ISO, 2002a). It is appropriate to collect soil samples according to soil depth in many industrial, commercial, residential, or urban soils where the surface soil horizons have been mixed or disturbed either through construction, landfilling, dumping or capping. Sample depths can be:

- standard depths (e.g., 0 to 15 cm, 0 to 30 cm, 15 to 30 cm, 15 to 60 cm, and 30 to 60 cm) (Mason, 1992; CCME, 1993a; Dalpé and Hamel, 2008);
- to the known vertical depth of contamination (e.g., 0 to 5 cm for soil contaminated through atmospheric deposition);
- ecologically relevant depths (those occupied by most soil organisms, usually 0 to 10 cm; no deeper than 30 cm) (Spurgeon *et al.*, 2002; ISO, 2006c);
- down to a known rooting depth (Dalpé and Hamel, 2008); or,
- consistent with sampling methods used in earlier phases of the site assessment or ancillary studies.

Regardless of what depth(s) are sampled, they must be consistent with the study objectives and meet the DQOs.

In addition, there are some challenges associated with sampling forest topsoil (*Ah horizon*) as it can sometimes be difficult to identify the exact boundary between overlying organic matter (forest floor) and the mineral horizon. For comparison of certain soil characteristics (e.g., soil carbon pools) among forest management practices, the collection of soil samples by depth may be preferred (Bélanger and Van Rees 2008).

If soil samples are collected at depth from uniform parent material, then it is advantageous to sample that material by depth increments to ensure that possible contaminants are not diluted. During sample design, soil and contaminant characteristics and expected contaminant transport mechanisms must be taken into account when selecting both sample depths and the range of depth that sample may include. This decision-making process will help ensure that contaminant concentrations within the soil are not diluted by selecting a sample depth range that is too large (e.g., sample collected from 1.0 m to 2.0 m).

When sampling soil at sites underlain by shallow rocky till with fluctuating topography, sampling at a consistent depth can become a challenge because soil depth can extend from only a few centimetres in one area to more than 1 m in another, depending on the sample location. When sampling in mountainous or hilly regions that have significant slopes (10° or more), the depth sampled needs to be corrected for the degree of slope using the *cosine rule*[27] with an extension factor of 1/cosine of slope (ISO, 2002a). The following example describes how to do this:

> A sample is to be taken to a 20-cm depth on a hill with a slope of 16°. To correct for the slope the sampling depth should be extended by a factor of 1/cos(slope), which in this case is 1/cos(16). Therefore to determine the corrected sample depth multiply the depth of the sample (20 cm) by 1/cos(16) = 20 cm × 1/0.96 = 20.8 cm (a difference of 4%).

Depending on the study objective, the LFH or O organic layers (if existing) are either removed before sampling according to depth (0 cm is the surface of the mineral soil) or is incorporated into the soil sample (0 cm is the surface of the organic layer). In both cases any surface vegetation and debris is removed prior to soil sampling (surface vegetation might also be

[27] The cosine rule is $c^2 = a^2 + b^2 - 2ab[\cos(C)]$, where a, b and c are lengths and A, B and C are angles, of a right angled triangle.

collected for analysis but it should not be included as part of the soil sample).

3.6.5 Collecting Soil Samples

The word "sample" is used within this document with the understanding that the sampling equipment collects a sample such that an unbiased estimate of the response being measured is possible. Factors that can affect the representativeness of the sample are:

- the distribution of particle sizes relative to the minimum aperture of the sampling equipment [a sampling device with a diameter of 2 cm should not be used to sample material that has aggregates > 2 cm diameter (e.g., gravel or pebbles), otherwise the sample will only contain smaller aggregate sizes and be biased]; this consideration determines the minimum size of a sample that can be considered representative of the (sub)population(s);
- aggregation of the contaminants of interest;
- non-homogenous distribution of contaminants within a sample; for example, many contaminants are associated with a particular size fraction.

Pitard (1993) discusses particulate sampling theory which underlies these ideas. ISO (2007c) provides guidance on collecting samples that are unbiased from these perspectives. If there is doubt regarding the lack of bias due to sampling equipment and/or methodologies, dialogue between statisticians and soil geophysicists may be required.

Shovels, scoops, or trowels are among the most commonly used tools in soil sampling when large volumes of soil are needed, and knives can also be used to slice through horizons; however, care must be exercised to ensure that a representative (see 3.3.5) and unbiased sample is collected (e.g., a constant depth or soil horizon must be removed). More precise sampling devices are soil corers, ring samplers, cutting frames, or soil cylinders but they are less convenient for extracting large soil sample volumes. However, note that soil cores or ring samples are most easily (and sometimes only) obtainable under moist field conditions (ISO, 2003a). Very large sample volumes are most easily collected using large earth-moving equipment such as a backhoe. Soil sampling devices should be selected in consideration of the following:

- objective of the study
- soil characteristics (sandy, clayey, very dry, very moist, frozen)
- location of contamination (surface versus at depth)
- properties of contaminant (volatile versus non-volatile)
- ecological receptors of concern
- biological tests to be conducted

These considerations will in turn help decide the best devices to use for:

- the type of soil to be collected;
- the depth(s) or horizon(s) to be sampled;
- the volume of soil to be sampled; and,
- the collection of disturbed or undisturbed samples.

Descriptions of the more commonly used soil collection devices as well as their advantages and disadvantages is provided in Table G.1 (Appendix G) and photographs and diagrams of select sampling devices are provided in Figures G.1 to G.6 (Appendix G). The list of sampling devices in Table G.1 is not exhaustive; there are many sampling devices available that are specific for different types of soil textures, conditions, volumes, and depths. Local reputable field sampling suppliers should be able to provide a wide range of devices as well as recommendations as to their suitability to meet specific study objectives.

3.6.5.1 Collecting Soils for Toxicity Testing
To collect soils:

1. Establish the boundaries of the sample location as per guidance in Subsection 3.3.8 (Figure 3).

2. Clear the soil surface of loose materials (i.e., debris, vegetation, or fresh leaf litter) by hand or by gentle raking.

3. If extensive vegetation covers the surface of the plot (e.g., thatch) then cut the vegetation to the surface of the soil where the sample is to be collected (Figure 4). When the soil sample is removed (follow steps 4 to 6), gently tap the soil from the roots into the sample container and either discard the plant material, keep it with the soil sample, or collect it as a separate sample depending on the study objectives (USEPA, 2006).

4. If large (e.g., > 1 L) sample volumes are required, extract the soil using shovels or trowels as follows:

   i. extract the soil from the top soil layer (the A mineral horizon) in scoops using trowels or a shovels to the depth of colour change (Figures 5 and 6).

   ii. place extracted soil into a sample container(s) (Figure 7). If soil is to be prepared on-site (i.e., dried, homogenized or sieved before shipping the sample to the laboratory), place the soil sampled onto a plastic sheet, cotton sheet (if contaminant(s) are plastic related), tarp or a *receiving container*(s) until the entire sample is collected and ready for preparation (Figure 8).

   iii. ensure that all the soil from the horizon or depth is removed from the entire plot

   iv. repeat steps 4(i) and 4(iii) for underlying B and *C horizons* if contaminated and/or depending on the study objectives.

5. If smaller sample volumes are required, or if consolidated samples (e.g., intact soil cores) are required, extract the soil using corers as follows (adapted from Mason, 1992 and ISO, 2003a):

   i. corers with wide diameters are recommended (e.g., 9 to 30 cm)

   ii. level the surface of the soil if necessary

   iii. drive the corer (using a mallet, steel, or slide hammer with a nylon head) to the known depth of the horizon or to the colour change that indicates the horizon transition (Figure 9) [a "test" core might have to be extracted from the plot first to determine where the colour change occurs in very heterogeneous sites (Figure 10)] or to the desired depth; if a split-core sampler is used (i.e., a sampler where the coring chamber can be opened after sample extraction and the core observed without disturbing it), the sampler can be driven to a standard depth and the extracted core can be removed and separated by horizons and/or depths (Figure 11).

   iv. avoid tilting the corer while pushing it into the soil

   v. extract the sample; minimize the loss of cored soil during extraction (e.g., create suction in the top of the corer barrel with a plunger or larger rubber stopper, or hand excavate around the outside of the core barrel to allow access to secure the bottom of the corer barrel prior to withdrawal).

   vi. cut away any transitional soil or roots and remove large stones from the bottom of the sample

   vii. for unconsolidated soil samples: extract the soil from the corer (or if a split-core sampler is used separate the sample by horizon and/or depth and then extract the soil from the sampler) and place soil into the sample container

   viii. for consolidated soil samples (intact soil cores): cap the cylinder, core liner

or sample ring and place it into a container for storage and shipment (Figure 12).

ix. for both unconsolidated and consolidated core samples, the depth of material retained should be recorded along with the depth of penetration of the core barrel to allow for an estimate of soil compression or loss during coring; it also provides information essential to subsequently relating biological test results to the original depth and thickness of the soil layers sampled.

x. if soil is to be prepared on-site (e.g., dried, homogenized, or sieved before shipping the sample to the laboratory), place the extracted soil onto a plastic sheet, cotton sheet (if cross-contamination of the sample from organic contaminants leaching from the plastic sheet is of concern refer to Subsection 3.3.11.1), tarp, or a receiving container(s) until the entire sample is collected and ready for preparation.

xi. repeat steps 5 (iii) to 5 (x) until all the soil required is extracted from the plot and placed in the same sample/receiving container

xii. to extract soil from subsurface horizons (B or C if necessary) repeat steps 5 (iii) to 5(x) either using the same locations, or new locations

xiii. where smaller volumes of soil are required, consolidated or unconsolidated samples can also be collected from the walls of pits (e.g., test pits), that have been excavated on-site. Samples are extracted horizontally from the pit wall using the same tools as those extracted vertically from horizons or depths with the added advantage that contamination while extracting the core from overlying horizons or depths is minimized.

6. If soil samples are collected at depth, extracting samples using a drill truck can be a more efficient and less labour-intensive method (Figure 13). Although these methods have been developed for extracting soil samples for the purposes of chemical or physical soil characterization they can also be used for collecting soil for biological testing. However, some drilling apparatus and techniques have the potential to modify the soil matrix and introduce artefacts that can influence biological test results. Air rotary drills, solid-stem *auger* techniques, and those that use various drilling fluids are not recommended; however, standardized methods published by the American Society for Testing and Materials (ASTM) (ASTM 2008b,c,d; 2009a,b) exist and might be appropriate depending on the study objectives and/or the DQOs.

7. If very large sample volumes are required (e.g., 100s of litres) then a backhoe is more efficient for sample extraction (Figure 14). Care must be taken to ensure that the soil is only removed to the appropriate depth or horizon.

8. For guidance on preparing soils on-site, refer to Subsection 3.6.6

3.6.5.2 Collecting Soils for Microbial Testing

All of the guidance in the previous subsection (3.6.5.1) also applies to collecting samples for microbial testing. However, there are a few extra considerations to make when sampling soil for microbial testing. When field conditions allow, soil should be collected with field moisture content that facilitates sieving; and, sampling should not be conducted during, or immediately following, long (e.g., > 30 days) periods of freezing, flooding, or drought (ISO, 2006a). In addition, it is very important to remove surface vegetation, roots, plant litter and soil fauna to reduce the addition of fresh organic carbon to the soil that can cause unpredictable changes in the structure and activity of the native microbial community (ISO, 2006a).

Figure 3. Demarcation of a sampling location (photo: D. Bright).

Figure 4. Trimming extensive vegetation to the soil surface at a sampling location prior to sample collection (photo: J. Princz).

Figure 5. Collection of a bulk soil sample at a site located in a river floodplain (photo: J. Princz).

Figure 6. Collection of a bulk soil sample in agricultural soil (photo: J. Princz).

Figure 7. Sampling soil within a boreal forest using 20-L polyethylene pails (lined with polyethylene bags) as sample containers (photo: G. Stephenson).

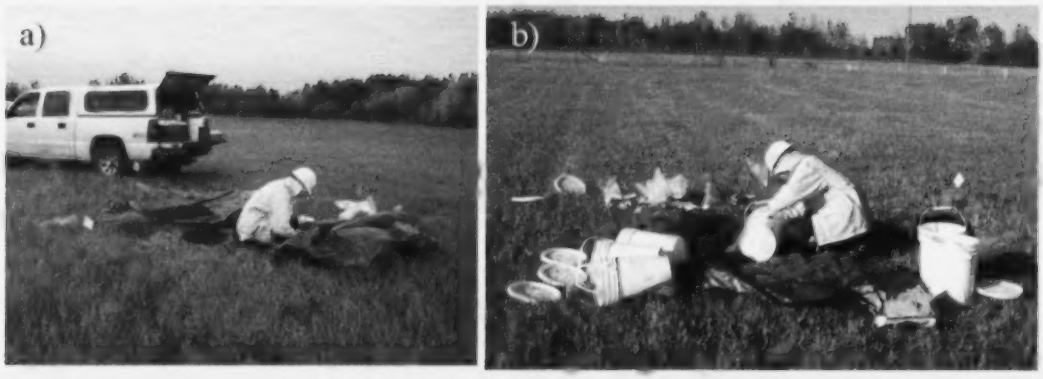


Figure 8. Temporary storage of bulk soil sample on a plastic tarp (a) until the entire sample is extracted and placed into sample containers (b) (photos: K. Bessie and N. Harckham).

Figure 9. Collection of a soil core using a slide hammer (EC and SRC, 2007).

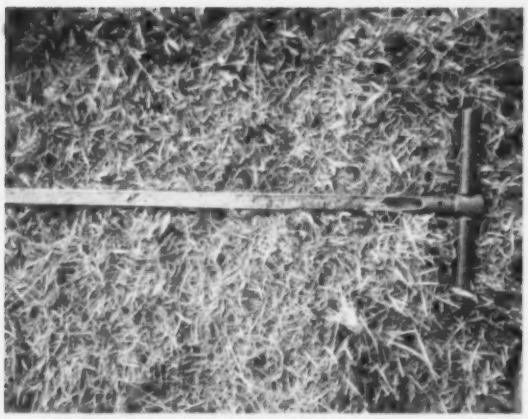


Figure 10. Example of a "test" core extracted using a Pirkhauer corer that can be used to characterize the soil profile (photo: J. Römbke).

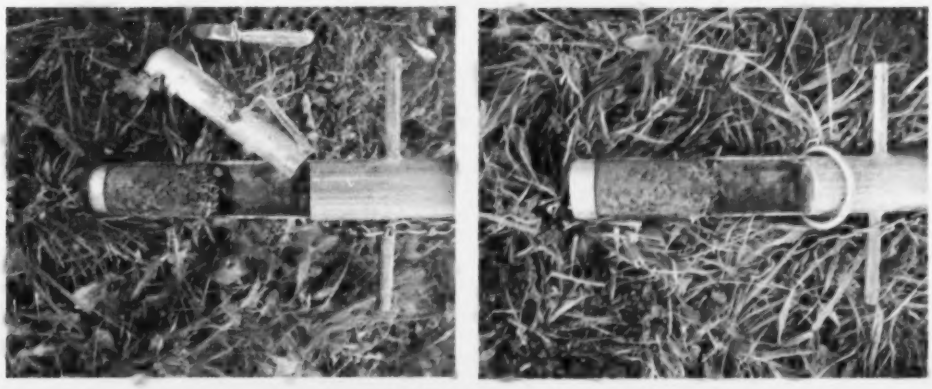


Figure 11. Soil core collected using a split-core sampler. Note that the coring chamber can be opened after the sample is extracted and the core can be observed without disturbing it (photos: J. Römbke).

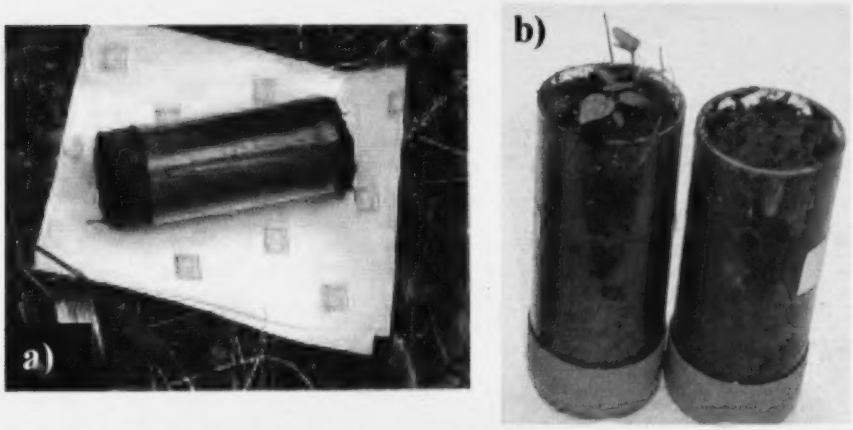


Figure 12. Soil core extracted from sampler (a) and contained intact within a plastic sleeve for use as an intact soil core test unit in a site-specific biological test (b) (Environment Canada and SRC, 2007).

Figure 13. Soil samples collected using a hollow stem auger with a drill (photo: K. Bessie and N. Harckham).

Figure 14. Collection of very large sample volumes of soil using a backhoe (photo: G. Stephenson).

3.6.6 On-Site Handling of Soil Samples

In some cases it is desirable to begin the preparation of samples for biological testing before the samples are shipped to the laboratory. On-site preparation might include hand sorting to remove debris and/or organisms, air-drying, sieving and homogenization of soil samples, prior to their placement into sample containers.

On-site handling might be necessary when:

- soil is very cohesive (e.g., clayey) and large chunks must be broken up and/or sieved (to make samples more homogeneous) (Subsections 3.10.3.3 and 3.10.3.4)
- soil contains very large rocks or stones and hand sorting or sieving ensures sufficient soil sample volume for testing (Subsection 3.10.3.3)
- soil contains large amounts of plant material, thatch, or roots and hand sorting or sieving ensures sufficient soil sample volume for testing (Subsection 3.10.3.3)
- soil is very wet (air-drying prevents the formation of anoxic conditions, reduces sample mass, reduces the potential for sample leakage and facilitates sieving) (Subsection 3.10.3.1)
- it is cost effective for field personnel to air-dry and sieve samples when appropriate conditions and equipment are present in the field

Air-drying should be conducted in the field only when necessary; it is preferable to air-dry soil under controlled, laboratory conditions (ISO, 2006a).

Soil samples collected for DNA or RNA analyses should be frozen in the field using dry ice (ISO, 2006a). Soil samples collected for microbial testing should be handled carefully and should be prepared as soon as possible after sampling. On-site preparation should include:

- removal of vegetation, larger soil fauna, and stones
- sieving through a 2- or 5-mm sieve (Subsection 3.10.3.3)

The outside of sample containers should be wiped clean after filling with soil to avoid cross-sample contamination. Any handling of samples undertaken in the field must avoid loss of sample, contamination from contact with other samples or soils from the site, and be consistent with the study objectives and the DQOs.

3.7 Sample Containers

The selection of the size, type, and composition of soil sample containers should be discussed with the toxicity laboratory(ies) and analytical laboratory(ies) (if samples for chemical analyses are to be sampled concurrently with samples for toxicity testing) during the planning stages of the study. The selection of the most appropriate sample container primarily depends on:

- the requirements of the biological tests to be conducted
- the volume of soil required for testing
- the type and nature of the soil contaminant(s)

Consolidated soil sample containers:
If the samples collected are to be tested as consolidated cores, they will be sealed in plastic sample liners and/or steel sample tubes. These plastic sample liners or sample tubes constitute the sample containers. Once all samples are collected, they should be secured in a larger container that will to keep them at specific

storage and transport conditions (e.g., cooled) and prevent them from being damaged during shipment.

Unconsolidated soil sample containers should be:

- large enough to hold the required soil volume
- made of inert material to minimize cross-contamination or reaction with, or adsorption of, soil contaminants
- strong enough to support the weight of soil and to protect sample integrity during storage and transport
- able to be sealed tightly to prevent spillage
- air-tight and pressure resistant if soil is contaminated with volatile compounds
- practical for handling in the field
- clean

All sample containers should be labelled in the field with the following information (see also Table H.1; Appendix H):

- sampling date
- sampling time
- sample identification number
- sample location
- sampling site
- sampler's name
- sampling conditions
- sample type
- sample volume

As an alternative to including all of this information on the sample container label (especially if there is insufficient room, such as would be the case for small samples), the sample container may be labelled with a unique sample identification number or name that is linked to a form that provides all other pertinent information.

Labels must be secured to each sample container; if a label is affixed to a container lid, an identical sample must be also secured to the container itself. Ensure that the information on the label is written with indelible ink and will not wipe off under transport or storage conditions.

All sample containers should be filled (as much as possible) to the top of the bucket to minimize free air space, unless expansion of the sample is anticipated (e.g., clay samples exposed to moisture). Containers that are recommended for soil samples collected for biological testing are listed in Table H.1; Appendix H.

3.8 Sample Transport

All samples should be stored until transport in the dark, protected from the environment (e.g., rain, wind), away from extreme heat and not be allowed to flood, dry out, or freeze. They should also be transported under those same basic conditions. Samples should be properly labelled and securely packaged for transport, especially if the sample containers are fragile. Large sample containers can be secured together with thin plastic wrapping into groups that can be stacked securely on skids for transport.

Smaller samples can be shipped conveniently by air or ground transport; air cargo transport is recommended if transport time needs to be kept to a minimum. Large soil samples (e.g., 20-L) usually have to be shipped by ground transport (truck, bus, or train). If large volumes of samples need to be transported under cooled conditions, some transport companies offer refrigerated vehicles for use; if not, these vehicles can be rented. Other considerations for sample transport follow.

- If samples contain contaminants that are stable and unlikely to undergo changes in bioavailability or speciation in the time between sampling and testing (e.g., weathered organics and inorganics), they can be transported under ambient, dark conditions.
- If samples contain volatile organic compounds or unstable contaminants (e.g., readily biodegradable or chemically reactive compounds), they should be transported under cooled (e.g., $4 \pm 2^{\circ}C$) conditions and with the minimum transport time possible.

- Samples for microbial testing should be kept dark with free access of air (if aerobic tests are to be conducted), kept cool (e.g., 4 ± 2°C), and compaction should be avoided (ISO, 2006a).

- Samples for DNA/RNA analyses should be frozen using dry ice and transported frozen to the laboratory in coolers using dry ice (ISO, 2006a).

- Consolidated soil samples (soil cores) should be packaged and transported with care to ensure that the original soil structure is not disrupted during transport (ISO, 2002b).

All samples must be shipped with appropriate documentation. This includes chain-of-custody (COC) forms as well as any specific regulatory documentation for transport of contaminated materials. Insurance should be acquired if necessary for reimbursement in the case of loss or destruction of the sample. The COC is a form used to track samples from the time of sample collection to the time of sample receipt at the laboratory (see example in appendix I). The COC should include (adapted from USEPA, 1986, 2006):

- site information (address, contact person, telephone number)

- client information (name and contact details) or project number

- sample identification number or code

- date and time of sample collection

- sample volume or mass

- brief description of the sample including contaminants of concern

- testing requested

- special instructions for soil handling, preparation or testing, including whether or not subsamples are to be collected for chemical analyses

- special safety precautions

- transport conditions

- printed names and signatures of all persons receiving and relinquishing samples (including couriers), and time and date of all sample transfers

- name and telephone number of transporting company (e.g., courier)

The COC is signed each time the samples change custody, and copies are kept by both the individuals relinquishing and accepting the samples. Particular care must be taken to document any change of custody when the sample is a *legal sample* to be used as legal evidence (e.g., to establish compliance with an environmental regulation). Documentation must be sufficient to prove that the integrity of the sample was not compromised from the time it was collected to the time it was tested. During this time, the sample must be under the physical custody of the person currently listed on the COC or protected under lock and key. Every individual responsible for sample custody must provide signatory documentation each time a sample is received or relinquished (USEPA, 2006).

If the concentration of the contaminants in the soil samples is high enough to render the soil hazardous material under the federal Department of Transportation (Transport Canada) *Transportation of Dangerous Goods (TDG) Act*, then all documentation according to the TDG regulations must be filled out and attached to the samples (http://www.tc.gc.ca/eng/tdg/menu.htm). In addition, samples must be labelled according to TDG regulations and appropriate placards must be displayed on any vehicle transporting the soil samples. If a third party is used to transport the samples (e.g., a transportation company), the site investigator must ensure that they operate under the TDG Act and follow the regulation (this step should be completed while developing the study plan). The same guidance applies to all provincial or local laws that apply to the shipment of dangerous goods.

3.9 Sample Receipt and Storage

3.9.1 Sample Receipt

When samples are received at the toxicity testing laboratory, all appropriate chain-of-custody (COC) documentation should be filled out immediately and a copy of the COC provided back to the sampler. Once received, the samples should be logged immediately according to relevant laboratory standard operating procedures (SOP) and QA/QC procedures. As a minimum, a unique sample identification code or number should be assigned to each sample, and the date received, sample type (e.g., soil), volume, and condition of the sample should be recorded.

Laboratory personnel should immediately inspect each sample and record the following in a sample receipt log:

- sample temperature;
- informal description of moisture content, texture and structure;
- presence of standing water;
- presence of indigenous invertebrates, fungi, or plant material; and,
- any strong odours, either chemical (e.g., petroleum) or natural (e.g., rotting material or indication of anaerobic conditions).

This information will aid the laboratory personnel in deciding which sample preparation procedures might be required before testing (if not pre-specified). Any special health and safety protocols should be reviewed as well as any special storage or handling instructions provided by sampling personnel or the contaminated site project manager, ideally prior to sample arrival.

3.9.2 Sample Storage

Sample storage conditions and duration is dependent on the nature of the soil contaminants, on the needs of the biological tests, and on the requirements of the test methods under which the biological testing is conducted.[28] The effect of storage time and temperature on soil toxicity and soil properties depends on the contaminant(s), soil characteristics, storage time, and conditions, especially temperature (Ehrlichmann *et al.*, 1997). Therefore specific storage conditions and durations should be discussed with the laboratory in the planning stages of the investigation and should take into consideration the study objectives, DQOs, soil conditions, and contamination.

All preparation, manipulation (if required) and testing should be conducted as soon as possible upon sample receipt. If samples contain volatile organic compounds or unstable contaminants (e.g., readily biodegradable or chemically reactive compounds), they should be stored under cool (e.g., $4 \pm 2^{\circ}C$) dark conditions and testing initiated as soon as possible. However, if samples contain contaminants that are stable and unlikely to undergo changes in bioavailability or speciation in the time between sampling and testing (e.g., weathered organics and inorganics), they can be stored longer (e.g., > 6 weeks) at room temperature (e.g., $20 \pm 3^{\circ}C$). In addition:

- samples should be prevented from freezing or partially freezing during storage (unless the samples were frozen upon collection)
- samples should be stored in airtight containers to prevent changes in water content during storage
- soil pH changes over time in storage (changes in pH can be rapid under certain conditions); if the magnitude of the change in pH is expected to unacceptably influence test results, then samples should be tested as soon as possible (Slattery and Burnett, 1992; Prodromou and Pavlatou-Ve, 2006)
- if samples have not been previously dried upon arrival, they should be stored at cool temperatures (e.g., $4 \pm 2^{\circ}C$) and/or air-dried and sieved as soon as possible (subsections 3.10.3.1 and 3.10.3.3); this is especially important for moist soils containing high amounts of organic material in order to suppress the growth and degradative

[28] For the Environment Canada earthworm, plant, and collembola test methods (EPS 1/RM/43, 45, and 47, respectively), it is recommended that soil toxicity testing should begin within two weeks of sampling and must begin within six weeks of sampling, unless it is known that the soil contaminants are aged and/or weathered and therefore considered stable.

activities of indigenous soil fungi and/or bacteria; profuse growth of fungi and/or bacteria in a soil sample has the potential to significantly affect the performance of test organisms in that sample

- if samples contain light-sensitive contaminants or soil constituents, they must be stored in darkness or in light-absorbent containers (ISO, 2002a)
- soil electrical conductivity, pH, macronutrients (P, K, Mg, and Ca), chloride, sulphate, fluoride, cyanide, and sulphide, non-volatile organics and trace metals in air-dried samples have been demonstrated to be stable following 6 months to at least 7 years of storage at room temperature (ISO, 2007a)
- nitrogen and its different organic and inorganic forms can fluctuate as the biological activity in the soil sample fluctuates in response to changes in aeration, temperature, and humidity; if nitrogen mineralization or ammonium oxidation is to be tested, or soil fertility is expected to significantly influence test results, then samples should be stored frozen if the microbial (see Subsection 3.9.2.1) or toxicity tests cannot be conducted immediately (e.g., within 4 days or less) (ISO, 2007a; Drury *et al*., 2008a)
- air-dried reference soils can be stored at room temperature for an indefinite period of time (ISO, 2007a), although there can be significant changes in nitrogen mineralization over time (Sheppard and Addison, 2008) as well as changes in soil pH

3.9.2.1 Storage for Soils Intended for Microbial Testing

There are special storage requirements and durations for soil samples intended for microbiological testing (ISO, 2006a):

- samples should be tested as soon as possible after sampling (e.g., 4 to 7 days) , if this is not possible, samples should be prepared before storage (Subsection 3.10.2)
- samples should be stored at 4°C and storage should not exceed 3 months unless there is evidence showing continued microbial activity
- samples should be stored in darkness (to avoid growth of algae) with free access to air (to avoid the development of anaerobic conditions); a loosely tied polyethylene bag is considered an appropriate storage container once the sample has been received
- samples should not be stacked, nor should the samples be too large as anaerobic conditions might develop at the bottom of the sample
- samples must not dry out or become waterlogged during storage
- if samples must be stored for longer than three months, they can be stored frozen at -20°C for up to 1 year; however, this is not generally recommended as freezing soil can cause significant and long-term changes in microbial abundance and activity (ISO, 2007a; Sheppard and Addison, 2008).
- samples that are to be tested for DNA/RNA or enzyme activity should be tested immediately; if this is not possible then:
  - samples for DNA and phospholipids fatty acid analyses, and dehydrogenase activity analyses can be stored at -20°C for 1 to 2 years
  - samples for RNA analyses can be stored at -80°C for 1 to 2 years after an initial shock-freezing with liquid nitrogen

3.10 Preparation of Soil Samples

The extent and type of soil preparation is dependent on the type of biological testing to be conducted, soil type, and the condition of the soil sample upon receipt at the laboratory. Figure 19 schematically illustrates the decision process of soil preparation from receipt of the soil sample to commencement of testing.

3.10.1 Preparation of Consolidated Samples for Tests with Intact Cores

Most biological tests (including single-species plant and invertebrate toxicity tests) can be conducted with both unconsolidated soil samples and consolidated samples. The test endpoints

measured are often the same, though the experimental design (e.g., number of replicates, number of organisms per replicate, volume of soil per test unit) can be quite different. Once consolidated soil samples (soil cores) are received at the laboratory, they should be stored immediately under cool conditions (e.g., 4 ± 2°C) and remain undisturbed until tested. Consolidated soil samples should be tested as soon as possible upon sample receipt and should not be frozen because freezing and thawing will disrupt the soil structure in the core and strongly influence microbial activity. By definition, the handling and preparation of consolidated soil samples should be kept to a minimum as it is desirable to keep the soil particles and pore structure similar to field conditions.

For single-species toxicity tests conducted with consolidated soil samples (e.g., intact cores), the extent of preparation for testing will depend upon the study objectives; however, procedures might include the removal of indigenous plant material from the surface of the core, and a slight disturbance of the surface of the core to facilitate the insertion of test plant seeds (Moody, 2006; Environment Canada and SRC, 2007).

3.10.2 Preparation of Unconsolidated Samples for Microbial Testing

This section provides recommendations for preparing unconsolidated soil samples to be used in microbial tests, including microcosm tests with soil functional, biomass, or diversity endpoints.

After soil samples have been properly logged at the testing laboratory, they should be prepared as soon as possible. Recommended procedures have been drawn from ISO (2006a) and Sheppard and Addison (2008):

- remove vegetation, larger soil fauna, and stones (if not already removed in the field)
- sieve the soil through a 2-mm sieve (maintains aerobic soil conditions)
- if soil is very organic (e.g., peat) and cannot be sieved at field moisture (moisture of the soil sample as it was received from the field) through a 2-mm mesh sieve, then a 5-mm mesh sieve should be used
- if the field moisture of the soil is too wet to sieve, the soil should be air-dried at ambient temperature to a "workable" moisture content (see Subsection 3.10.3.1).
- avoid air-drying if possible as drying-wetting events can induce significant changes in microbial C and N dynamics; drying-rewetting procedures, commonly used when preparing soils and then pre-incubating soils by adding moisture, can significantly influence respiration rates, substance utilization efficiency, nitrification potential, microbial biomass, nitrate and ammonium concentrations of soil samples, and these changes can persist for more than a month following the stress (Fierer and Schimel, 2002); re-wetting after drying also causes bursts of respiration and growth of distinct populations of bacteria (ISO, 2006a); therefore testing soils as soon as possible after sample collection, and thereby avoiding the necessity of drying and rewetting, is recommended; however, if the field moisture of the soil is too wet to sieve, the soil should be air-dried at ambient temperature to a "workable" moisture content (Subsection 3.10.3.1)
- when drying the soil it should be spread out to facilitate uniform drying, and should be finger crumbled and turned frequently to avoid excessive drying
- do not air-dry soil for DNA analyses (Subsection 3.10.3.1) as it will change the microbial composition (Topp *et al.*, 2008)
- homogenize soil samples thoroughly after sieving (Bailey *et al.*, 2008)
- soil samples should not be ground or crushed (Drury *et al.*, 2008b)
- intact cores, rather than bulk samples, are recommended for evaluating nitrification since, at least in some soils, slight impacts in soil structure can significantly increase the levels of extractable nitrate (Drury *et al.*, 2008a).

Once the soil samples are prepared, or have been removed from storage, they should be pre-incubated to allow germination (and removal) of seeds and to re-establish the microbial metabolic equilibrium following the change from sampling or storage conditions (ISO, 2006a; Sheppard and Addison, 2008). Pre-incubation conditions should be as similar to the test conditions as possible. Pre-incubation durations are dependent on the test, soil type, and storage conditions; durations that range from 2- to 28-d are considered acceptable (ISO, 2006a).

3.10.3 Preparation of Unconsolidated Soil Samples for Toxicity Testing

The section describes recommended procedures to prepare unconsolidated samples for testing other than microbial testing. Users of this guidance should bear in mind that any soil preparation has the potential to profoundly affect the biological test results. As with soil storage, there is no one way that is suitable for all soil types, contaminants, study objectives, and soil conditions. There are some more commonly practiced methods; however, and these will be discussed in the following subsections. The goal of soil preparation is to minimize disturbance of sample conditions and as much as possible while making the soil amenable for testing and not interfering with the study objectives and DQOs. As with soil sampling and storage procedures, proper documentation should be made of any soil preparation procedures conducted.

3.10.3.1 Drying

Drying of soil samples might be required if the sample:

- is too moist to homogenize
- is too moist to sieve (if sieving is required)
- contains standing water
- is very moist and anaerobic conditions are present in the sample (unless this is a desired test condition)
- is very moist with high organic matter content and cannot be tested immediately (in order to suppress profuse growth of indigenous fungi and bacteria)

Do not dry soil samples if:

- the method or study objectives requires soils to be tested at field moisture content

Soils should be air-dried at ambient temperature to a "workable" moisture content. A "workable" moisture content is a soil moisture content usually between field moist and completely air-dried soil (completely air-dried soil has moisture in equilibrium with the humidity of the ambient air, or "soil dried right out"). A "workable" moisture content results in soil sufficiently dry to allow sieving without the production of dust and/or permits the soil to flow freely to facilitate thorough homogenization. Soils should not be completely air-dried. Soils should not be oven-dried at all. Completely air-drying soil, or oven-drying soil to any degree, causes a number of reactions in the soil: microfauna and flora die or become senescent; contaminants dissolved in soil pore water become concentrated, precipitate or coagulate; and, solid organic matter deforms, can become hydrophobic and/or expose underlying mineral surfaces (Sheppard and Addison, 2008).

To air-dry soils:

- spread out the sample on a sheet of thick polyethylene or polypropylene plastic or clean stainless steel surface and air-dry the soil only until it is dry enough to sieve/homogenize (soil is "workable"); leave sufficient margins between the soil and the edge of the work surface to avoid contamination from other sources
- the time required to dry might range from 1 hour to 5 days, depending on the moisture content of the soil, the depth to which the soil is spread out, the texture of the soil, and the temperature and humidity of the room
- ensure that while drying the soil, the sample is placed in a location safe from disturbance, dust and/or contamination
- if the soil sample has a high clay content, ensure that it does not dry out (e.g., air-dry completely); when clay soil completely dries it becomes very cohesive, hard and difficult to handle; monitor clayey soil samples frequently to prevent them from becoming

too dry (e.g., push small amounts of soil through a sieve frequently while drying)

- if the soil sample is saturated upon arrival at the laboratory (Figure 15a) (e.g., presence of standing water, which is common if soils were collected frozen or at a site with a high water table), then spread the soil out either on a thick plastic sheet (Figure 15c) or a clean flat container with sufficient surface area to allow drying. Incorporate the standing water into the spread-out soil sample if possible
- if there is too much standing water to incorporate it all into the sample at one time:
  - decant the standing water before spreading out the soil (Figure 15b).
  - store the decanted water in a glass container under cool (e.g., 4 ± 2 °C) conditions
  - dry and then sieve the soil as per the guidance in this subsection (3.10.3.3)
  - spread the soil out again on a plastic sheet/stainless steel surface and incorporate as much of the decanted water as possible
  - dry the soil again—do not sieve—homogenize the soil as per Subsection 3.10.3.5; incorporate as much of the remaining decanted water as possible.
  - repeat the previous step until all of the decanted water has been incorporated into the soil and then homogenize the soil again

Soils are usually subject to wetting and drying cycles in the field; therefore, air-drying soils to a workable moisture content (e.g., a moisture content more representative to that found in the field) is a common, and a commonly considered defensible, soil preparation technique (for toxicity testing, see earlier cautionary comments regarding this practice for microbiological testing) (Sheppard and Addison, 2008).

3.10.3.2 Wetting

Hydrating soil samples might be required if the:

- sample is too dry to prepare for testing
- sample is so dry that it poses a human health hazard through inhalation of small particulate matter upon handling or testing

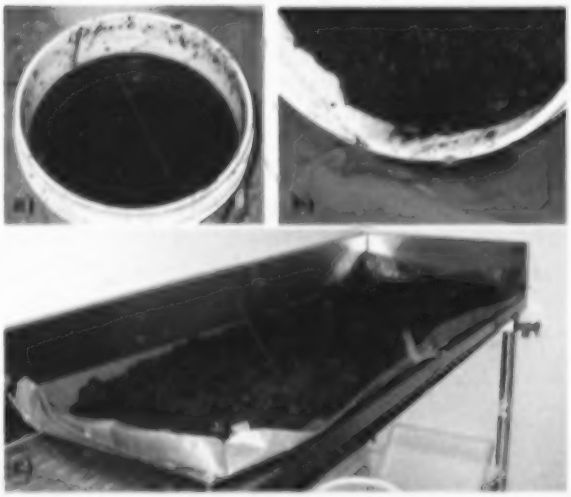


Figure 15. Saturated soil sample a) received as is; b) being decanted of excess water; c) air-dried to a workable soil moisture content on a clean plastic sheet in a laboratory (photos: J. Princz).

Soil samples should be hydrated with purified (e.g., de-ionized, reverse osmosis) water only, to avoid introducing cations and anions (including Cl^-) or trace metals (e.g., Cu^{2+}) into the soil. Incorporation of these tap water constituents cause *artifacts* of soil preparation and can change soil electrical conductivity, soil pH, or metal concentrations, which in turn can influence test organism performance. Re-wetting must be done slowly and in stages; soil moisture hysteresis might result in a lower water-holding capacity than expected and care must be taken that soils do not become too wet for further preparation or testing.

Soil samples can be hydrated by:

- spraying purified water over the surface of a sample that has been spread out over a flat surface and then homogenizing the sample by hand on the flat surface, or by hand or machine in a bowl or pail

- by gently pouring purified water over the surface of the spread-out soil sample in small increments and then homogenizing the sample by hand on the flat surface, or by hand or machine in a bowl or pail (Subsection 3.10.3.5)
- spraying purified water over the surface of the sample that has been placed into a test unit
- adding the purified water to the dry soil to test units slowly in small increments with a pipette
- adding purified water to soil in test units by laying a piece of filter paper on top of the soil and then hydrating the soil by saturating the filter paper (recommended for wetting soils with a very fine texture to avoid developing a hardpan over the surface of the soil in the test unit)

After soils have been wetted, they should be left to equilibrate with the purified water at least 24 hours prior to testing.

3.10.3.3 Sieving

Sieving of soil samples might be required if the:

- sample contains undesirable or large amounts of thatch, plant roots, or visible indigenous flora or fauna
- sample contains stones or rocks too large for testing (e.g., ≥ 6 mm diameter)
- sample aggregation is heterogeneous (e.g., large and small clumps in sample)
- sample texture or organic matter distribution is visibly heterogeneous
- soil aggregates are > 6 to 10 mm
- sample structure is not amenable for testing (e.g., sample consists of clay chunks from surface or subsurface soil)

Do not sieve soil samples if:

- the samples were already sieved in the field (confirm with field personnel)
- the samples have a crumbly texture that is optimal for testing (EC, 2004a, 2005a, 2007a)

Soils are typically dried-sieved using a 2-, 4-, 6-, 8- or 10-mm mesh (Zagury *et al.*, 2002; Sheppard and Addison, 2008; ISO, 2006a). Sieving through a 2-mm mesh sieve is standard as the ≤ 2-mm size class encompasses all of the soil separates or fractions (sand, silt, and clay); however, in some cases, this can be restrictive for the purposes of biological testing. The goal of soil preparation is to keep the disturbance of the sample to a minimum; sieving through a 2-mm mesh usually requires the soil to be dried to a lower workable moisture content than if soil is sieved through a larger mesh size. Soil sample aggregation should be maintained as closely as possible to that in the field; more aggregates of a larger diameter are preserved when using a larger mesh size. In addition, soil contaminants can be associated with particles or objects in the soil > 2 mm in size; removing these particles might bias the sample by reducing soil contamination (Mason, 1992). This is particularly important when contamination is present in the form of large particles of inconsistent sizes (e.g., explosive material residues) or when contaminants (e.g., metals) are closely associated with plant roots or other plant material > 2 mm. Conversely, larger particles kept in the sample might result in dilution of the contamination if it is associated with particles with ≤ 2 mm diameter. The rationale for sieving soil through mesh sizes between 4 and 10 mm is that it represents a compromise between maintaining sample integrity as much as possible and enabling the soil to be thoroughly homogenized and used in biological tests. The mesh size used for sieving depends on the soil type, contamination, organic matter content and type, and moisture content, and must meet the study objectives and DQOs. Soil should be sieved carefully to minimize loss of soil sample and dust generation. Sieves with a stainless steel mesh should be used.

Soil samples comprised of clayey subsurface soils can be received as large chunks of moist, very cohesive soils (shovelfuls of clay) (Figure 16a). It is not unusual for ecotoxicity testing to be conducted on very clayey soils due to the contamination of subsurface clay soils as a result of oil and gas drilling activities in Canada. Moist, clay samples are very cohesive and

cannot be sieved or homogenized; therefore,to prepare this type of sample:

- spread out the sample on a sheet of plastic as previously described
- manually break apart soil aggregates to approximately 0.5- to 2-cm aggregates (Figure 16c), dry the soil to a workable moisture content as described in Subsection 3.10.3.1
- sieve (Figure 16d) (subsection 3.10.3.3) and/or homogenize (Subsection 3.10.3.5) the soil depending on the study objectives

Manually breaking the soil into 0.5- to 2-cm aggregates is extremely time-consuming and labour intensive, but it avoids the necessity of having to air-dry and crush or grind the soil. Air-drying and crushing or grinding the soil is more efficient and cost-effective and provides a more homogenous sample; however, it is not suitable for clayey soils contaminated with volatile organic compounds.

In some cases, grinding clayey subsurface soils received as large chunks of dry clay is preferable to manually breaking apart soil aggregates and then sieving; refer to subsection 3.10.3.4 for recommended preparation techniques.

Figure 16. Highly cohesive clayey subsurface soil samples a) received as is; b) manually broken into approximately 5-cm aggregates; c) manually broken into 0.5- to 2-cm aggregates; and, d) sieved to ≤ 0.6-cm aggregates (photo: G. Stephenson).

3.10.3.4 Crushing or Grinding

Grinding, crushing or milling soil samples might be required if:

- the sample structure is very difficult to homogenize by sieving (e.g., clayey soil)
- greater homogeneity of the sample is desired than can be achieved by sieving
- greater homogenization of thatch or organic material into the sample is desired than can be achieved by sieving.

In general, grinding of soil samples is to be avoided when possible. Grinding is a very aggressive preparation technique that completely destroys the soil structure and aggregation. It also changes soil surface area, microbial mineralization, nutrient availability, and contaminant bioavailability to a greater extent than sieving (Sheppard and Addison, 2008). However, there are some circumstances where grinding is desired, or even necessary, in order to test the soils.

If organic soil, or organic material and thatch in mineral soil samples cannot be sieved, even using sieves with large (e.g., 10 mm) mesh sizes, these samples might be processed successfully by grinding them with knife mills (Sheppard and Addison, 2008). To ensure that grinding or milling does not introduce metals to the soil sample, laboratory equipment should be manufactured from carbide steel. This should be confirmed with the analytical laboratory prior to requesting analysis.

As mentioned in Subsection 3.10.3.3, the preparation of soil samples with high clay content can be especially challenging. These soils are often received as large chunks of very dry hard clay that cannot be sieved or homogenized (Figure 16a). This soil can be prepared by either crushing or grinding.
To crush the soil:

- spread the soil out on polyethylene or polypropylene plastic or a stainless steel surface

- manually break up any large chunks of clay that separate relatively easily (Figure 16b)
- allow the soil sample to dry until it is completely air-dry
- using a hammer or mallet, break the chunks into aggregates that are approximately 0.5 to 2 cm in diameter (Figure 17); the sample is then ready for homogenization (Subsection 3.10.3.5)

To grind the soil:

- spread the soil out on polyethylene or polypropylene plastic or a stainless steel surface
- manually break up any large chunks of clay that separate relatively easily (Figure 16b)
- allow the soil sample to dry until it is completely air-dry
- using a hammer or mallet, break the chunks into aggregates approximately 5 to 7 cm in diameter (Figures 17, 18a)
- grind the dried clay chunks in a mechanized clay grinder (Figure 18b); the finely ground sample (Figure 18c) is then ready for homogenization (Subsection 3.10.3.5)

The preferred method to prepare dry clayey soil samples is to crush the soil rather than to grind it, as crushing is less aggressive than grinding. In addition, when soils are crushed they retain an aggregate structure that makes them easier to hydrate when biological tests are set up. Ground soils typically have a powder-like consistency and can be difficult to re-hydrate unless it is done in stages and with caution. However, crushing the soil is more labour-intensive and time-consuming. Grinding the soil is more efficient and can provide a significantly more homogenous sample than crushing the soil. The choice to crush or grind a soil sample depends upon the study objectives and soil contaminants and should be discussed among all investigators at the early stages of the study.

3.10.3.5 Homogenization

Homogenization of soil samples might be required if the:

- sample is heterogeneous and reduction in this heterogeneity is desired
- sample is a composite sample
- sample is an unconsolidated sample

In general, all unconsolidated soil samples should be homogenized at the laboratory prior to testing, even if the sample properties and contamination are considered homogeneous. This is because unconsolidated samples tend to settle out into different particle fractions during transportation from the field, especially non-cohesive, very dry soils (e.g., dry sandy soils) (ISO, 2002a). Particles become distributed by size; smaller particles migrate to the bottom of the container while larger particles remain at the top. This spatial segregation by particle size also occurs in storage but to a much lesser extent; therefore, it is recommended that samples be homogenized just prior to testing (Mason, 1992).

There are a number of methods used to homogenize soil samples, including:

Folding (ISO, 2007a)

- appropriate for small sample volumes (0.5 to 10 L)
- spread the soil in a thin layer over a sheet of thin plastic
- fold the layer with the plastic and spread the soil out again
- repeat several times

Mixing

- appropriate for medium to large sample volumes (20 to 60 L)
- empty the soil in the sample container in the middle of a sheet of thick polyethylene or polypropylene plastic or stainless steel surface into a conical pile

- spread the pile out on the sheet to a depth no higher than the height of the mixing tool
- divide the soil into four equal quarters using a shovel or other rigid tool
- mix each quarter manually, or with a trowel or rake or other appropriate tool for the same period of time (e.g., 2 to 5 minutes)
- place the entire sample into a large flat container with ledges, lips, or sides so thorough mixing can occur without sample spilling or spreading out
- homogenize entire sample by mixing thoroughly by hand or rake or other appropriate tool for the same prescribed time (e.g., 2 to 5 minutes)

Coning (adapted from ISO, 2006d)

- appropriate for medium to large sample volumes (20 to 100 L)
- empty the soil in the sample container in the middle of a sheet of thick polyethylene or polypropylene plastic or stainless steel surface into a conical pile
- using a scoop or shovel, systematically remove soil from the base of the cone and form a second cone with all the material from the first cone transferred to the apex of the second cone
- repeat the coning twice

Figure 17. Manually crushing air-dried cohesive clayey subsurface soil to approximately 0.5- to 2-cm aggregates (photo: G. Stephenson).

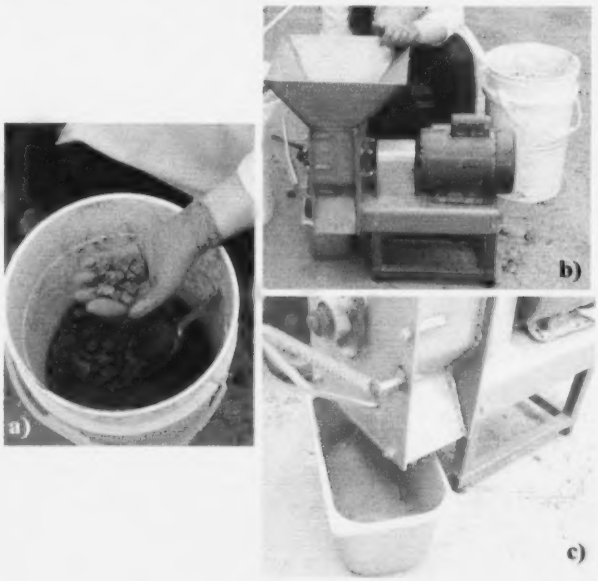


Figure 18. Grinding very cohesive clayey soils by a) manually separating soil into 5- to 7-cm aggregates and air-drying completely followed by b) grinding in a stainless steel clay grinder to produce c) a very finely ground soil sample (photo: G. Stephenson).

3.10.3.6 Reconstitution

Reconstitution of constituents of soil samples might be required if the:

- standing water was decanted from the surface of the soil sample during preparation
- portions (e.g., thatch) of the sample were removed during preparation but need to be tested along with the soil
- soil horizons collected as separate soil samples are tested as layers re-stratified within a single test unit

If standing water was decanted from the surface of the soil before soil preparation began, it should have been stored under cooled conditions until the sample was dried and sieved. The soil should be then reconstituted with this water as described in subsection 3.10.3.1.

If soil contamination is associated with organic soil horizons [e.g., the *LFH horizon* of forest soils, or the organic surface layer (O horizon)] of cryosols from northern Canada, then it is prudent to test soils with all of the organic material present. However, these layers of soil can contain significant amounts of thatch, plant roots, and other organic material that might be removed from the soil samples by sieving. In order to incorporate the sieved plant material back into the soil sample the sieved material must first be ground (e.g., using a knife mill subsection 3.10.3.4), pulverized (e.g., using commercial or domestic chopping machines) or chopped (manually or mechanically using commercial or domestic equipment). Once the soil sample is dried, sieved, and homogenized then the processed organic material can be reincorporated into the sample and re-homogenized according to subsection 3.10.3.5.

Contaminated soil sampled for biological testing should by default be collected and tested by soil horizon (ISO, 2002a, 2006a; EC, 2003a). If the soil contamination extends beyond a single horizon, then re-constructing the different soil horizons in a test unit provides an ecologically relevant test system and can generate data on contaminant distribution, bioavailability, and toxicity among horizons as well as test organism preference or avoidance of horizons. Using this approach can also potentially reduce the volume of soil collected per horizon. Reducing the volume of soil collected can result in significant savings in time and labour as the volume requirements for bulk soil testing can be quite high. The depth to which horizon is reconstructed in a test unit is study-specific. For example:

- if surface horizon depths in the field are very shallow, as occurs in some forest soils in Canada, it might be possible to layer the topmost horizons (e.g., FH or A horizons) to the actual field depths
- the depths of each layer in the test unit could correspond to the relative horizon depths observed in the field
- the depths of each layer could be operationally set to meet specific study objectives [to maximize exposure of test organisms to all soil horizons (Moody, 2006; EC and SRC, 2007)]

To construct a test unit with multiple horizons, each horizon is prepared individually (dried, sieved, homogenized, and then prepared according to the procedures outlined in a standard biological test method), and then placed into a test unit according to volume or mass. Each subsequent layer is placed on the previous layer carefully so as to avoid inadvertent horizon mixing. This multi-horizon layered test method is currently under development by Environment Canada, but shows promise for the biological testing of highly stratified soils using plants (EC and SRC, 2007). Unfortunately this method is not applicable for invertebrate test species. Invertebrate species will move down in the soil to the lower portion of the test vessel regardless of the horizon (EC, 2009).

It should be noted that this is an appropriate approach to prepare soil for testing that was stratified at the site at the time of sampling. In some situations, the site owner or manager may have taken preliminary remedial action and disturbed or mixed the natural soil horizons. In these scenarios the soil sample will have to be

tested as a mixed soil without an attempt to re-layer the different horizons for testing.

3.11 Physical and Chemical Characterization of Soils

In order to better interpret biological test results, and to be in compliance with some (e.g., Environment Canada) test methods, it is recommended that the following suite of physical and chemical characteristics be analyzed for each contaminated soil sample:

- particle size distribution (percentages of clay, sand, and silt)
- total organic carbon[29]
- organic matter content[29]
- pH
- electrical conductivity
- water-holding capacity
- nitrogen as total N, nitrate (NO_3^-), nitrite (NO_2^-) and ammonium (NH_4^+)
- potassium as total and/or plant-available
- phosphorus as total and/or plant-available
- C:N ratio
- cation exchange capacity
- major anions and cations (Na^+, K^+, Mg^{2+}, Ca^{2+}, Al, S^{2-}, Cl^-)
- oxidation-reduction potential
- soluble salts
- sodium adsorption ratio (SAR)
- contaminants of concern
- characteristics of the contamination (e.g., odour, staining, debris, presence of fuel or solvent).

Ensure that the analytical methods used are consistent with, or specified by the study DQOs. Carter and Gregorich (2008) and Carter (1993) provide guidance on a variety of soil analyses.

As described in Subsection 3.6.2, each reference soil sample should be analyzed for the following:

- particle size distribution (percentages of clay, sand, and silt)
- organic matter content[29]
- pH
- electrical conductivity
- fertility
  - nitrogen as total N, nitrate (NO_3^-), nitrite (NO_2^-) and ammonium (NH_4^+)
  - potassium as plant-available or total potassium
  - phosphorus as plant-available and total phosphorus
  - C:N ratio (for microbial tests)

In addition, other properties to measure in reference soil might include:

- cation exchange capacity
- total carbon content[29]
- total inorganic carbon[29]
- exchangeable cations (Ca^{2+}, Mg^{2+}, Na^+, K^+)
- oxidation-reduction (redox) potential
- water-holding capacity

As described in Subsection 3.6.2, in order to confirm that the reference soils are not contaminated, the following screening analyses are recommended:

- organophosphorus insecticide suite
- organochlorine insecticide suite
- herbicide suite
- metals suite
- petroleum hydrocarbons (including PAHs)
- other site- or area-specific contaminants of concern

It is also strongly recommended that the soil concentrations of the contaminants of concern be measured, as a minimum, at the beginning of the toxicity test in order to confirm the exposure concentrations. Exposure concentrations might be different than expected from field analyses

[29] Organic matter content can be used to calculate total organic carbon (TOC) by multiplying the organic matter content of a soil by a soil constant (USEPA, 2002c). However, the relationship between TOC and OM is slightly different among soils and the total organic carbon content should also be determined by laboratory analysis.

due to variability of contamination at the site, sample collection procedures or sample preparation procedures. It is particularly important to measure the exposure concentrations of volatile and unstable contaminants at the beginning of a biological test (see Section 4). Exposure concentrations should also be measured at the end of a test and at least one point in time (preferably three points in time) during the test so that any degradation of contaminants, and corresponding decreased exposure concentrations, can be documented.

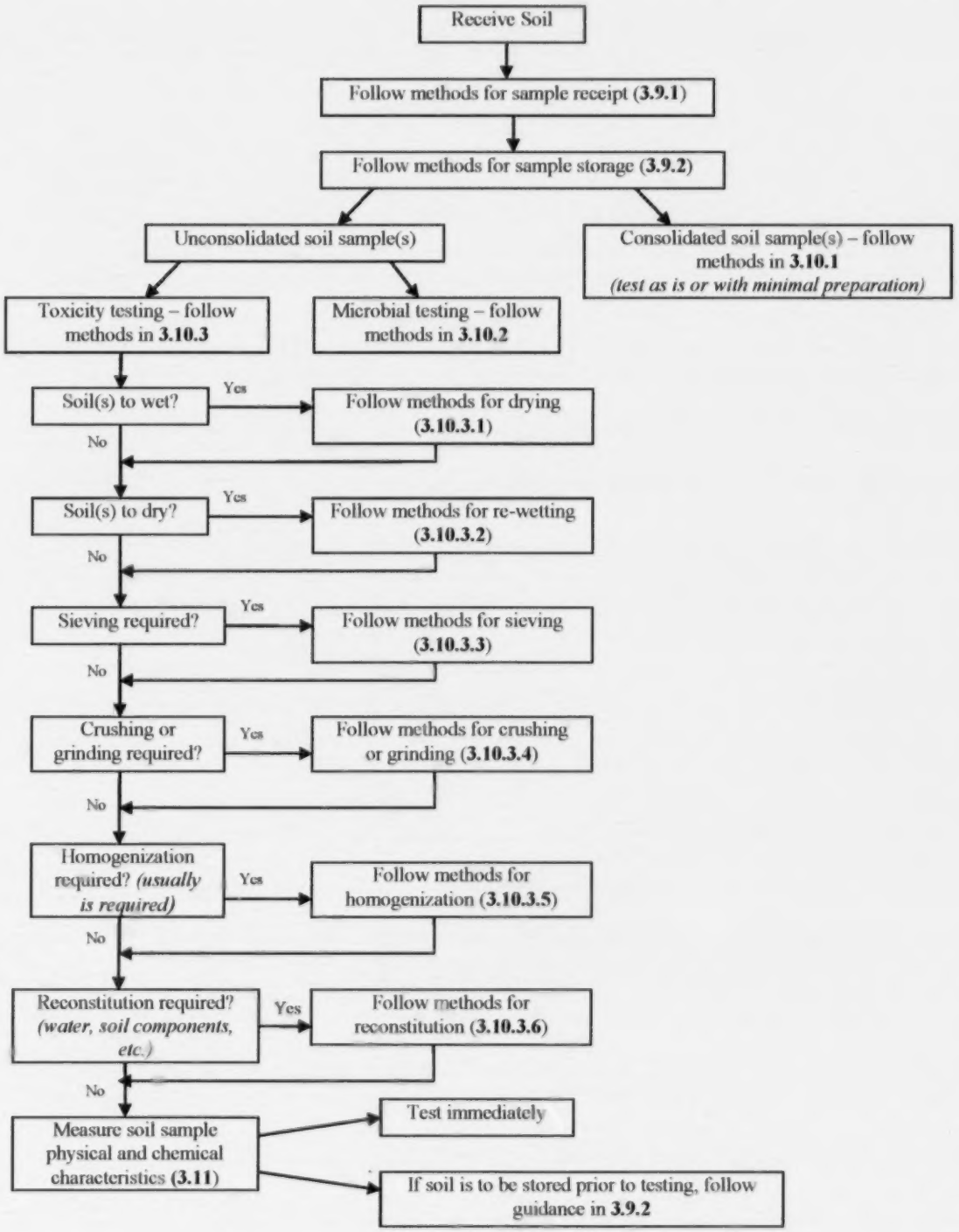


Figure 19. Summary of soil preparation process (corresponding subsections highlighted in boldface).

Specific Procedures for Collection, Handling, and Preparation of Soils Contaminated with Volatile or Unstable Compounds

The primary concern when testing soils contaminated with volatile, semi-volatile, and/or unstable contaminants is to minimize the loss of these contaminants when sampling and handling the soils in the field, transporting soils to the toxicity laboratory, and, once at the toxicity testing laboratory, to minimize loss of these contaminants prior to testing (e.g., during sample storage, handling, or preparation).

4.1 General Guidance

When samples are contaminated with volatile, semi-volatile, and unstable compounds, the following general guidance applies:

- minimize the time between soil sampling and biological testing;
- minimize the loss of volatiles or degradation of compounds during sampling, transport, storage, preparation, and testing;
- maintain conditions that reduce the potential for volatilization (e.g., cool 4 ± 2°C) or degradation (e.g., keep samples in the dark, keep samples cool to minimize microbial activity);
- measure the concentrations of volatile and readily degradable compounds frequently so that the magnitude of loss is known as well as where the most loss occurred;
- measure the concentrations of volatile and readily degradable compounds just prior to biological testing in order to have an accurate measurement of the initial exposure concentrations in the test; and,
- measure the concentrations of the volatile and readily degradable compounds throughout the duration of the test.

The following procedures are not appropriate when samples are contaminated with volatile, semi-volatile, and unstable compounds; samples should not be: composited, air- or oven-dried, ground or crushed, irradiated or autoclaved; and, samples should not be amended with compounds that inhibit microbial activity.

4.2 Sample Collection

In order to minimize the loss of volatiles or degradation of compounds during sampling:

- consolidated samples (soil cores) should be collected using a core liner or tube so that the samples can be sealed following extraction from the soil
- if samples are collected without the use of core liners, extract the soil from the sampler as quickly as possible and place into a non-permeable sample container (e.g., glass or plastic sample jars, or stainless steel containers)
- soil cores should be sealed with Teflon caps or wrapped in Teflon tape and then coated with a non-reactive sealant
- once soil cores are sealed they should be placed in a cooler with ice packs to keep the samples as cool as possible
- if the study objectives permit, samples could be frozen using dry ice[30]
- samples should be kept away from light to minimize photo-degradation
- if samples are contaminated with weathered semi-volatile compounds, unconsolidated small or bulk samples can be collected; however, collection should be carried out quickly, analytical subsamples should be

[30] When shipping with dry-ice, samples should be packaged and labelled according to federal *Transportation of Dangerous Goods (TDG) Act* requirements.

taken, and samples should be sealed, cooled, and transported immediately

- unconsolidated sample containers should be filled completely to minimize headspace unless the sample is clayey; clay samples can swell in transit when moist and break glass or plastic sample containers

4.3 Sample Transport

To minimize the loss of volatiles or degradation of compounds during transport:

- samples should be shipped cool (e.g., 4 ± 2°C) in a cooler or a cooled transport vehicle
- soils destined for microbiological testing should not be frozen, unless they were collected that way
- unconsolidated samples should be sealed and air-tight; stainless-steel buckets with push-fit lids can be used for large volume (e.g., 10-L) samples
- samples should never sit on a loading dock exposed to ambient conditions

To minimize the time between soil sampling and biological testing:

- samples should be shipped to the toxicity and analytical laboratories using the quickest method possible; usually via air transport with a courier service
- make sure that sampling does not occur just before a weekend or a holiday, otherwise samples may sit in a warehouse without temperature control until the laboratory opens
- notify laboratory personnel about the shipment so they are prepared to process the samples immediately upon arrival (and obtain confirmation of shipment receipt and cooled conditions upon receipt of the samples from the toxicity laboratory personnel)
- testing should commence as soon as possible

4.4 Sample Storage

To minimize the loss of volatiles or degradation of compounds during storage:

- samples should be stored under cool (e.g., 4 ± 2°C) dark conditions for a minimum period of time

4.5 Sample Preparation

To minimize the loss of volatile compounds during preparation of unconsolidated samples:

- remove the sample containing volatile compounds from cool storage
- keep preparation to a minimum, and prepare the sample as quickly as possible
- if the sample must be dried, dry the soil in the fume hood according to Subsection 3.10.3.1 and minimize the drying time as much as possible
- if the sample must be sieved, sieve the soil in the fume hood as quickly as possible according to Subsection 3.10.3.3 with the following more detailed recommendations:
  - keep unsieved soil in a sealed pail
  - sieve one handful of soil onto a plastic sheet under the sieve; when the handful of soil is sieved, put it into the sealed pail containing sieved soil
  - do not sieve soil into an open pail; keep the pail containing sieved soil sealed until ready to receive the next handful of sieved soil
- if the sample must be homogenized, homogenize in the fume hood as per Subsection 3.10.3.5 as quickly as possible
- when soil preparation is finished, continue to store the sample in a sealed container under cooled conditions until just prior to testing

These preparation procedures, recommended for unconsolidated samples, will result in significant loss of highly volatile compounds, and some loss of semi-volatile compounds. If it is critical that the concentrations of volatile compounds in the soil remain very similar to those in the soil collected in the field, the only appropriate procedure is to test sealed consolidated samples (sealed intact soil cores). No preparation or

manipulation is conducted with consolidated samples and the following pre-test procedures are recommended:

- on the day of test setup, consolidated samples are removed from cool storage and placed under a fume hood
- the sample is unsealed and as quickly as possible the sample is placed into the test unit and the test unit closed
- test organisms, when added, are added in as short a time as possible to prevent loss of volatile compounds

To minimize the loss of readily degradable compounds during preparation:

- follow the same procedures described for consolidated or unconsolidated samples; however, it is not necessary to conduct these procedures under a fume hood

To minimize the loss of photolabile compounds during preparation:

- samples do not necessarily need to be prepared in the fume hood
- follow the same procedures described for consolidated or unconsolidated samples; however, conduct these procedures under minimal lighting conditions; for unconsolidated samples, follow these additional procedures:
  - when drying soil, dry it in a dark room or in a closed cabinet; ensure that the room or cabinet can only be opened by appropriate personnel;
  - do not cover soil with any material while it is in the dark room or cabinet as it will impede the drying process;
  - when sieving soil, any time that the soil is removed from the storage container it should be covered with an opaque material (e.g., thick black garbage bags), even while sieving a handful through the sieve;
  - when homogenizing soil, cover the spread out soil with opaque material (e.g., thick black garbage bags) and sieve under the material (this might necessitate that one person homogenizes the soil while a second person holds the opaque material very closely over the soil so that there is sufficient room for the first person to homogenize)
- soil samples for chemical analyses should be collected in amber containers and stored in complete darkness

4.6 Sample Contaminant Analyses

The concentrations of volatile and readily degradable compounds should be measured frequently so that the magnitude of loss is known as well as where most of the loss occurred. As a minimum, measurements should be taken from analytical samples collected in the field, and from subsamples collected on the first day that test organisms are introduced into the test units (Day 0 of the test). These two measurements will provide the site investigator with the extent of the contaminant loss between sampling and test setup and whether there has been any preferential loss of contaminant constituents (e.g., a greater loss of the more volatile contaminant constituents). Measuring the concentrations of the contaminants in the soil at the beginning of a biological test is critical; this measurement represents the initial exposure concentration for the test organisms, which can be significantly lower than the concentrations in the field.

The following additional measurements are recommended to determine the magnitude, nature, and timing of loss of volatile, semi-volatile, or unstable contaminants from the time of sample receipt to the end of the biological test(s):

- concentrations of contaminants in the samples upon arrival at the laboratory
- concentration of contaminants following preparation of samples (e.g., drying, sieving, homogenization)
- concentration of contaminants at different points throughout the biological test and at the end of the test

Specific Procedures for Manipulation of Soils in Preparation for Testing

The section describes possible procedures that might be required for manipulating unconsolidated soil samples to render them testable to meet study objectives or DQOs when the conditions do not occur with the samples as collected. Soil manipulation is more disruptive to sample integrity than soil preparation; it is usually conducted with the intention of influencing soil contaminant bioavailability, toxicity, or soil organism performance. It is not generally recommended to manipulate soil physicochemical characteristics in order to accommodate test organism performance. Instead the preferred approach is to select organisms that are known to tolerate the physical and chemical conditions (e.g., soil pH, salinity, etc.) of a given soil sample(s). Therefore, soil manipulation should be considered only when necessary to meet study objectives when other alternate approaches are not possible. As a result, it is particularly important that detailed proper documentation should be made of any soil manipulation procedures conducted.

Manipulation of consolidated soil samples (e.g., intact soil cores) by definition is not appropriate or recommended. Manipulation of soil samples prior to microbiological testing is not recommended.

5.1 Washing

Objective of manipulation: to reduce soil electrical conductivity or salinity prior to testing.

Washing soil samples might be required if the:

- soil is too saline (e.g., elevated electrical conductivity) and the elevated salinity will confound toxicity test results [the soil has elevated salinity (naturally saline, or saline as a result of contamination)] and reduction in electrical conductivity is desired to differentiate between the effects of elevated salinity and of soil contamination on test organisms
- soil is too saline (naturally or as a result of contamination) to support survival or growth of test organisms; however, it is still desirable to estimate the toxicity of co-contaminants

Do not wash soil samples if:

- the samples are contaminated with water-soluble compounds unless purpose of washing is to remove water-soluble co-contaminants

To wash soils:

1. Measure the electrical conductivity (EC) from the samples following receipt or obtain EC values for the soil sample(s) from the site investigators.
2. In general:
   - EC values greater than approximately 1.5 dS/m can have adverse effects on earthworm survival or reproduction (Kerr and Stewart, 2003).
   - EC values greater than approximately 2.5 dS/m can have adverse effects on the emergence and growth of sensitive to moderately sensitive species (which include many standardized test species) (Blaylock, 1994).
   - Note that these values are approximations only; the sensitivity of the test organisms in any given study will dictate levels at which washing might be considered.
3. If EC values are considered high enough to potentially adversely affect test organism performance independent of other soil contaminants, then consider washing the soil sample.
4. If soils are to be washed, only sieve soil if an excessive amount of thatch or gravel is present.
5. Do not homogenize soil.

6. Separate soil sample into multiple 20-L clean pails (unlined) by filling each pail to approximately 1/3 of the volume.
7. Fill ½ of each pail with *de-ionized water*.
8. Stir the sample slowly into a slurry.
9. Allow the de-ionized water-saturated sample to settle for a minimum of 24 hours.
10. Decant the de-ionized water from the sample.
11. Discard decanted water.
12. Measure EC.
13. If EC levels are still too high, repeat steps 7 to 12 until EC levels are acceptable.
14. Once EC levels are acceptable, remove the soil slurry from the pails and spread the soil out either on a thick plastic sheet, stainless steel surface, or a clean flat container with sufficient surface area to allow drying.
15. Once soil is dry enough, sieve the soil to re-create soil structure as per Subsection 3.10.3.3.
16. Once sieved, homogenize the sample as per Subsection 3.10.3.5.
17. Because of the destructive nature of this soil preparation, it is strongly recommended to collect one or more subsamples of soil for analyses of the contaminant(s) of concern before and after washing to: measure loss of or changes to contaminants through manipulation and to measure the exposure concentrations at the beginning of the test(s).

5.2 Aging/Weathering

Objective of manipulation: to simulate the often mitigating effects of aging and/or weathering processes on contaminant bioavailability in the field.

Aging or weathering soil samples might be required if the:

- study objectives require contaminants to be weathered or aged; usually only applicable to soil samples amended with contaminants at the laboratory

Aging is a phenomenon that describes the change in contaminant bioavailability to soil-dwelling organisms with an increase in soil contact time. Bioavailability is generally decreased due to soil physical and chemical processes such as diffusion into soil nano-pores, and adsorption and desorption to soil organic matter (Stantec 2004; Loibner *et al.*, 2006; Jensen and Mesman, 2006). Strong sorption, slow release and limited diffusion processes result in the sequestration of hydrophobic contaminants and renders sequestered residuals recalcitrant to bioremediation (Jensen and Mesman, 2006). The processes governing aging of metals in soils includes incorporation into mineral structures, diffusion into pore spaces within minerals, nucleation/precipitation, mineral surface oxidation, and entrainment via the formation of chemical complexes within soil solids (Wendling *et al.*, 2009). Aging is usually a two-phase process; an early rapid sorption stage (e.g., to clay, soil humic acid, soil-derived metal oxides, etc.) followed by slower reaction processes (e.g., adsorption on condensed organic matter, diffusion into nano-pores, encapsulation) (Jensen and Mesman, 2006; Wendling *et al.*, 2009). Weathering is the relative change in the composition of contamination due to preferential loss of constituents with time (Stantec, 2004).

Ecotoxicity test data using metal- or organic-amended (spiked) soils often overestimate the bioavailable fraction of the contaminant in soils at a site because they do not account for the aging and weathering processes that have occurred at the contaminated site over time. One of the primary benefits of assessing the toxicity of soils collected from contaminated sites is that the contaminants in the soil samples are already aged and weathered by natural processes in the field. However, an investigator might wish to simulate the field bioavailability of a contaminant if, for specific study objectives, one or more soil samples collected from the site has been freshly amended with a contaminant or co-contaminant; or, the test results in the site soil is to be compared with test results in soil amended

with a contaminant or co-contaminant in the laboratory.

There are no standardized methods for aging and weathering; however, a few common approaches include:

- passive aging/weathering where amended soil is stored for short or long periods of time (at fixed or fluctuating temperatures) (Ma *et al.*, 2006)
- subjecting the amended soil to multiple wetting and drying cycles (Kuperman *et al.*, 2005, 2006) where soils are hydrated to a standard moisture content (e.g., 60% of the water holding capacity) and then allowed to air-dry for 7 days; repeat the cycle for up to three months, depending on the chemical and soil characteristics
- subjecting amended soil to multiple freezing and thawing cycles during which soils are frozen and then thawed at prescribed temperatures and durations, or at ambient (but documented) temperatures and durations

One or more subsamples of soil should be collected for analyses of the contaminant(s) of concern before and after aging/weathering to measure loss of or changes to contaminants through manipulation and to measure the exposure concentrations at the beginning of the test(s).

5.3 Adjusting Soil pH

Objective of manipulation: to increase or decrease soil pH prior to testing.

Adjustment of soil pH in samples might be required if the:

- effects of soil type and/or contaminant(s), independent of, or as influenced by soil pH, is under investigation (e.g., discriminate pH effects from contaminant effects)
- changes in pH are expected at study site (e.g., from acid rain or rehabilitation activities)

Adjusting soil pH is not appropriate:

- to improve the test performance of biological test species that are not tolerant of soil with naturally extreme (high or low) pH
- to adjust soil pH of either reference or contaminated soil, or both, when the pH of the soils are not closely matched

Many studies have investigated the significant influence of soil pH on contaminant bioavailability, especially metal contaminants. Soil pH influences metal speciation and also some soil organics. Metal bioavailability generally increases with decreasing soil pH (Sauvé, 1997; Dayton *et al.*, 2006; Echevarria *et al.*, 2006) though not for all metals and metalloids (Langdon *et al.*, 2003). Low soil pH itself can also cause significant stress to test organisms (Janssen *et al.*, 1997). Many soils in Canada have soil pH less than neutral (e.g., < 7) and some widespread soils have very low soil pH (< 4) such as podzols in the boreal shield ecozone. Low pH can also be a co-contaminant in soils as a result of industrial activities. Conversely, soils with extremely high pH (> 8) can also significantly influence biological test results. Very alkaline soils can also occur naturally (e.g., alkaline solonetzic soils in central Canada) or become so as a result of anthropogenic activities. Rather than manipulating the pH of the soil to accommodate the needs of the biological test organisms, test organisms should be selected based on the tolerance of the species to conditions (e.g., pH) of the soil and soil types found within ecozones from which it is collected. A discussion on the selection of appropriate test species according to soil found in the different ecozones of Canada is provided in Section 6.

Adjusting the pH of soil samples, however, might be considered if: the investigator is trying to determine the influence of soil pH on contaminant bioavailability; or soil conditioning (Subsection 5.4) significantly changed soil pH. Soil pH can be raised or lowered, depending on the study objectives and soil conditions. To adjust soil pH:

1. Prepare the soil sample(s) as per guidance in Subsection 3.10.3.

2. Measure the moisture content, pH, and electrical conductivity of the soil sample(s).

3. Choose the substance to add to increase or decrease soil pH.

   a. to increase pH:

      i. calcium carbonate ($CaCO_3$)[31], lime, gypsum, or other calcareous substances can be used

   b. to decrease pH:

      i. sulphuric acid (H_2SO_4), hydrochloric (HCl) acid, or any other strong (or weak) acid can be used

   c. the selection of the pH adjusting substance depends on the:

      i. individual laboratory experience
      ii. study objectives
      iii. soil contaminants
      iv. possibility of co-contamination by constituents of the pH-adjusting substance (e.g., HCl will increase the concentration of chloride ions and might become toxic to test organics if concentrations become sufficiently elevated)

4. Calculate the mass or volume of substance to be added to the soil on a dry weight basis to induce the desired pH change.

5. The amount of substance to be added should be pre-determined through preliminary experiments with a subsample of the soil to be tested; it is important that the preliminary experiments be conducted with the same soil to be tested as pH adjustment is soil-specific.

6. Incorporate the substance into the test soil using a liquid carrier if necessary (e.g., purified water).

7. Homogenize the soil thoroughly.

8. Incubate the soil (3 to 5 days).[32]

9. Measure the moisture content, pH, and electrical conductivity of the adjusted soil.

10. If soil pH needs to be further adjusted, repeat steps 3 to 9.

11. If soil pH is within an acceptable range of the target pH (this range should be predetermined and should be in accordance with study DQOs), hydrate the soil to the level required by the test method.

12. Measure the pH of the adjusted soil.

13. If soil pH is no longer within the acceptable target range, repeat steps 3 to 12.

14. If soil pH is still within the acceptable target range, initiate the test.

[31] Calcium carbonate is an effective substance to use for increasing soil pH. Calcium carbonate has the advantage over other calcareous-based substances in that it is readily available in a purified powder form, is not known to be toxic to most soil organisms, and the buffering capacity of calcium carbonate is greater than some alternatives such as lime or gypsum (e.g., raised pH is more stable over time) in some soil types (Stephenson, 2003). "Lime" can refer to a number of different compounds including calcium hydroxide and calcium oxide that have greater buffering capacity than calcium carbonate. However, these compounds are riskier to use because these compounds can raise soil pH much higher than calcium carbonate (up to pH 8) if over-used.

[32] Note that 3 to 5 days is an approximation only and depending on the soil buffering capacity, magnitude of pH change required, and agent used to modify pH, more or less than 3 to 5 days might be required and need to be taken into consideration when scheduling tests. This is another reason why conducting preliminary tests with subsamples of soil prior to manipulating the entire subsample for testing is recommended.

Preliminary investigations can also include assessments of how stable the soil pH changes are over time and to what extent pH is affected by the hydration of the soil to test requirements. This preliminary information can save much time and soil, and prevent frustration when pH-adjusting large volumes of soil for testing, especially when time and soil are limited. It is further recommended that when adding substances to increase or decrease pH, that the substances are added with a minimal amount of carrier (e.g., deionized water) as the soil can become more cohesive and aggregates can become larger every time moisture is added to the soil when it is re-homogenized. This is particularly true for very cohesive soils such as clays where homogenization itself can cause increased aggregation of soil. If pH-adjusted soil samples are part of a toxicity test, the resultant experimental design must always include treatments of unadjusted reference and contaminated soils for the purposes of comparison, and for QA/QC.

One or more subsamples of soil should be collected for analyses of the contaminant(s) of concern before and after pH adjustment to measure loss of or changes to contaminants through manipulation and to measure the exposure concentrations at the beginning of the test(s). The Environment Canada terrestrial toxicity test methods also provide guidance on how to work with soils that have extreme pH (EC, 2004a, 2005a, 2007a).

5.4 Conditioning Soils

Objective of manipulation: to change the structural or textural characteristics of soil prior to testing to improve soil physical conditions for test organism performance.

Conditioning soil samples might be required if:

- the unconditioned soil (in particular subsurface soil) represents a suboptimal habitat for test organisms to the extent that the influence of soil texture and/or structure is deemed to unacceptably confound biological test results

- if other study objectives require that the texture or structure of the soil be changed

Soil properties can vary widely among samples received for biological testing and some soils constitute a more acceptable habitat for the growth and/or reproduction of test organisms than others. Some soils, typically surface horizons with greater than 4% organic matter (OM), easily support growth and/or reproduction of the test species that are currently recommended in standard test methods. Soils that contain between 3% and 4% OM, and also have a medium soil texture (e.g., loams) do not generally pose a problem for standard test species. However, soils that have a low organic matter content (< 3%) and a very fine (e.g., clayey) or coarse (e.g., sandy) texture might constitute unsuitable habitats for plant growth or invertebrate reproduction, even when plant emergence and invertebrate survival is unaffected. Earthworm reproduction, in particular, can be adversely affected in sandy or clayey soils with low OM (Jänsch *et al.*, 2005).

As mentioned at the beginning of Section 5, the preferred approach is to use test organisms that are tolerant of the condition of the soil sample, in this case soil texture and organic matter content. However, the current suite of standard test species for Canadian soils (EC, 2004a, 2005a, 2007a) are limited to those species that perform optimally in soils with medium to low bulk density and OM greater than 3%. This is particularly true for most of the current standard earthworm test species (e.g., *Eisenia* spp.) (Stantec, 2008). When earthworms are tested with "suboptimal" soils reproduction is often adversely affected. For example, earthworm reproduction can be adversely affected when tested in both reference and contaminated clayey subsurface soils. However, because of the suboptimal environment, the extent of the impact (if any) on organism performance due to exposure to contaminants is unclear.

Until alternate earthworm (or other soft-bodied invertebrate) standard test species more appropriate for testing the "suboptimal" soils (as described) are available, one way to address this issue is to amend these soils with an appropriate soil conditioner. In the context of this guidance, a soil conditioner is an amendment that will reduce bulk density, optimize water retention, reduce compaction, and improve aeration in the

soil as well as be biologically inert (it will not be a food source for test organisms) and will not reduce the bioavailability of the contaminant(s) (it will not bind or sorb the contaminants). Soil conditioners can be organic or inorganic, and must be of consistent composition from batch to batch. Commercial products are recommended as they are more widely available and have consistent production processes.

A few substances have been evaluated that meet most or all of these criteria for soil conditioners for clayey and sandy soils with little organic matter (< 2%). They include *Sphagnum* peat, coconut coir (shredded coconut husk material) (Garcia, 2004), perlite, and gypsum that are added to soil on a dry-mass basis (Stantec, 2008). *Sphagnum* peat and coir were effective as conditioners. Earthworm reproduction has been shown to significantly increase in clayey and/or sandy soils amended with *Sphagnum* peat and coir at levels ranging between 2.5% and 15% (on a dry-mass basis) (Stantec, 2008). In other studies, the amendment of clayey soils with peat at 2.5% to 20% (dry mass) has also significantly improved earthworm reproduction and plant growth (Stantec, 2005, 2006). However, the drawbacks of these particular conditioners is that they do have the potential to decrease the bioavailability of organic (e.g., PHC) contaminants, and because of their natural acidity, the pH of the soil sample is lowered following amendment. The decrease in the pH caused by the amendment, however, can be overcome by adjusting the soil pH (Subsection 5.3). Sand is another conditioner that has been added to fine-textured soils with high bulk density to increase the porosity of the soil for test organisms (Zagury *et al.*, 2002). However, the addition of sand does not increase the OM content of the soil, which is important for earthworms and by increasing soil porosity, the greater amount of circulating soil water can act as an extractant and increase contaminant bioavailability (Zagury *et al.*, 2002).

Although there is a focussed research effort currently examining best practices for soil conditioning, at present there is no standard soil conditioner to use for "suboptimal" soils, nor can a standard level of amendment be recommended. The level of amendment must be assessed on a soil-specific basis; this is not surprising given the natural heterogeneity of soil properties. In addition, if the soil conditioner used decreases contaminant bioavailability, the level of amendment chosen will be a compromise between acceptable test organism performance and minimal decrease in bioavailability. The level of amendment chosen should be the lowest possible that will still achieve the study objectives. Regardless of the soil conditioner and level of amendment chosen, the resultant experimental design must always include treatments of unamended reference and contaminated soils for the purposes of comparison, and for QA/QC.

In addition to conditioning the soil for the purposes of improving test organism performance, sometimes study objectives require that the structure or texture of the soil be uniform among treatments. For example, if multiple soil samples are to be evaluated that are similar in contamination levels and all other characteristics except sand content, samples could be amended with a clean, standardized sand conditioner to a uniform level among treatments.

To amend soil samples with a soil conditioner:

- prepare the soil samples as necessary (Subsection 3.10.3)
- prepare the soil conditioners, if necessary, (e.g., sieving) using the same procedures used for soil samples
- measure the moisture content of the soil sample(s) and conditioner(s) once they are ready for testing
- amend the soil with the soil conditioner, both on a dry mass basis
- homogenize the soil + conditioner thoroughly
- measure soil pH and electrical conductivity of the amended soils
- adjust pH if necessary (Subsection 5.3)
- monitor pH of soil until desired pH is reached if necessary (Subsection 5.3)

- prepare amended and unamended soils according to the standard test method

One or more subsamples of soil should be collected for analyses of the contaminant(s) of concern before and after conditioning to measure loss of or changes to contaminants (including contaminant bioavailability) through manipulation and to measure the exposure concentrations at the beginning of the test(s).

5.5 Adjusting Soil Fertility

Objective of manipulation: to change soil fertility while minimally changing other soil characteristics prior to testing.

Adjustment of soil fertility in samples might be required if the:

- soil sample fertility is low and limits the performance of test organisms to an unacceptable extent
- soil fertility is unequal among soil samples and unacceptably confounds test results

The fertility of the soil can significantly influence the performance of test organisms. Although soil invertebrates are influenced by soil fertility indirectly through the amount of organic matter present in the soil; plants are influenced directly. Nitrogen is the major limiting nutrient in soils determining plant growth, followed by potassium and phosphorus, although any nutrient can be limiting if not available in sufficient quantities for plant growth (Bélanger *et al.*, 2008).

Contaminated soil can be nutrient-poor, particularly if the soil was collected from an industrial site and/or a heavily eroded site. The influence of the nutrient status of a site on test organism performance is important to evaluate as part of a site assessment and the fertility of site soils should not normally be manipulated. However, if plant growth at a contaminated site is very poor, and the nutrient status is also poor, the investigator might wish to determine if plant growth is inhibited due to the contaminants at the site, or is limited by the lack of essential nutrients. In addition, if the toxicity of a site soil is to be evaluated by comparing plant growth in the contaminated soil relative to plant growth in a reference soil, and the nutrient status is dissimilar between the two soils, it is prudent to adjust the fertility in one or both soils so they match. For example, if plant growth is better in the reference soil, and fertility is higher in the reference soil, it is unclear if the reduced growth in the site soil is due to contamination or to fewer nutrients. Conversely, if the fertility and plant growth is greater in the contaminated soil, it is unclear if the greater fertility in the contaminated soil compensated for the toxicity of the soil contamination or if the site soil was truly non-toxic.

To adjust the fertility of soil samples:

1. Prepare the soil samples as necessary (Subsection 3.10.3).

2. Measure the nutrient status of all relevant soil samples by submitting subsamples to an accredited soil nutrient analysis laboratory.

3. Soils should be analyzed for, as a minimum, the following:

    - total nitrogen and nitrogen as nitrate (NO_3^-), nitrite (NO_2^-) and ammonium (NH_4^+)
    - total and/or plant-available potassium
    - total and/or plant available phosphorus
    - calcium
    - magnesium

4. Analyzing micronutrients (e.g., Mn, Fe, Ni, Cu, Zn) should also be considered.

5. Once the nutrient status of the samples are known, calculations can be made:

    a. to equilibrate the nutrient status among soils for each nutrient (on a dry mass basis), or

    b. to determine whether or not fertility is suboptimal for the plant test species and which nutrients need to

be supplemented and to what degree[33]

6. To incorporate nutrients into the soil, it is easiest to formulate soil-specific nutrient solutions.

7. There are a wide variety of nutrient formulations available that can be used to formulate study-specific nutrient solutions; nutrient solutions can also be custom-made using nutrients obtained from laboratory chemical suppliers.

8. Guidance on the preparation of nutrient solutions are widely available; an old but still useful and commonly used reference is Hoagland and Arnon (1950).

9. Amend the soil(s) with the nutrient solution (EC, 2005a).

    - a second set of subsamples can be sent for analyses at this stage to confirm that the nutrient status of all relevant soils is equal
    - if more adjustments need to be made, repeat the process described above

10. Commence preparations for test setup.

One or more subsamples of soil should be collected for analyses of the contaminant(s) of concern before and after fertility adjustment to measure loss of or changes to contaminants through manipulation and to measure the exposure concentrations at the beginning of the test(s).

[33] Determining the optimal nutrient status of a soil is not always straightforward. Optimal concentrations of one or more nutrients are soil-specific, plant species-specific and are influenced by the concentrations of other macro- and micronutrients as well as soil pH. For example, zinc deficiency might be a result of low zinc levels in the soil, or might be caused by high levels of phosphorus in the soil that is inhibiting the uptake of zinc. Interpretation of soil fertility results and fertilizer recommendations might require consultation with, or analysis by, an agronomic laboratory and/or provincial agricultural extension personnel.

5.6 Reducing Indigenous Soil Microorganisms

Objective of manipulation: to reduce the indigenous microfloral and faunal populations in soil samples prior to testing. Indigenous microorganisms might need to be reduced if the:

- sterilized soil is used in tests estimating soil biomass (e.g., fumigation-extraction method)
- sterilized soil is used as control soil in soil enzyme assays, soil respiration tests, direct DNA extraction studies, for example
- test require axenic conditions
- soil contaminant is susceptible to rapid loss via microbial degradation, especially if soil samples cannot be tested immediately

The reduction of the indigenous microbial populations in soil samples is not a common or standardized practice. One of the reasons for this is because it is very difficult to effectively reduce these populations, and it is almost impossible to eradicate them (e.g., sterilize the soil), especially if it is necessary to treat large volumes of soil (Trevors, 1996). It is difficult to sterilize soils because microorganisms in a senescent life stage (e.g., cysts or spores) can often survive treatments and might in fact flourish following treatment due to reduced competition. In addition, because soil is a three-dimensional matrix with many microhabitats, most "sterilization" treatments cannot extend into the interior of the soil sample being treated, and/or cannot penetrate completely into these microsites.

Although it is impractical to attempt to sterilize large volumes of soils for testing, there are methods that can substantially reduce indigenous microbial populations. The two most efficient and widely used methods include repeated autoclaving and irradiation (Trevors, 1996), which are described in more detail in Subsections 5.6.1 and 5.6.2, respectively.

5.6.1 *Autoclaving Soils*

- soils are subjected to prolonged (e.g., minutes to hours) periods of extreme moist heat (e.g., > 100°C)
- autoclaving time should be increased with increasing volumes of soil to sterilize (e.g., 1 hour of autoclaving for any samples larger than 500 g (Trevors, 1996)
- to be effective, it is important that soils be autoclaved in thin (e.g., 1 cm) layers, as a result, only limited amounts of soil can be treated at a time
- soils should not be compacted but left uncompressed to allow for penetration of the steam
- better sterilization is achieved if soil is not saturated; soil air-dried to approximately 60% or less of the water-holding capacity of the soil is recommended if soil is saturated (Trevors, 1996)
- it is recommended that a soil sample be autoclaved three or more times to maximize the reduction of microorganisms (Lotrario *et al.*, 1995; Trevors, 1996; Sheremata *et al.*, 1997); soil samples should be incubated for 24 hours or more between autoclave cycles to allow spore germination (which will subsequently be destroyed in the next autoclave cycle)
- soil can be pre-incubated prior to the first autoclaving to reduce the number of autoclave cycles required for the maximum reduction of microorganisms; to do this, hydrate the soil to approximately 0.3 mL water/1 g soil and pre-incubate at ambient temperatures for 7 days prior to the first autoclaving cycle
- the extreme heat of autoclaving often induces chemical reactions in the soil; in particular, the organic acids can become transformed, organic contaminants can become transformed, and the physical and chemical interactions between contaminants and soil particles can be modified
- these modifications can result in an increase or decrease of soil bioavailability and toxicity (Lotrario *et al.*, 1995)
- once autoclaved, utilize test soils within one day of the last autoclaving cycle or store soils at -20°C or -80°C until use to suppress re-growth of senesced organisms that survived the autoclaving.

5.6.2 *Irradiating Soils*

- soils are subjected to prolonged (e.g., days, weeks) exposure to low levels of high frequency radiation (e.g., gamma radiation)
- gamma radiation from a cobalt-60 source is the most commonly used to irradiate soil and typical doses include 10, 25, or 50 kGray (kGy) gamma radiation (Lotrario *et al.*, 1995; Trevors, 1996; Sheremata *et al.*, 1997; Renoux *et al.*, 1999)
- other doses of radiation can also be used and Jackson *et al.* 1967 (as cited in Trevors, 1996) found that irradiating 30 g samples of soil with 1 Mrad (10 kGy) was required to kill all fungi and 2 to 3 Mrads (20 to 30 kGy) was required to eliminate all bacteria; up to 4 Mrad (40 kGy) have been used to sterilize soil
- individual soil subsamples exposed to the radiation source must be kept small (e.g., 250 g) to allow energy to penetrate to the centre of the subsamples; however, large numbers of subsamples can be exposed at one time
- soils are most easily exposed to radiation in polyethylene bags at the desired test moisture content
- one drawback to using irradiation is that the specialized facilities and expertise required are not widely available
- the high energy of the radiation induces chemical reactions in the soil (Trevors, 1996); Renoux *et al.*, 1999); therefore, as with autoclaving, the organic acids can become transformed, organic contaminants can become transformed, and the physical and chemical interactions between contaminants and soil particles can be modified
- these chemical reactions and modifications can result in an increase or decrease of soil bioavailability and toxicity (Renoux *et al.*, 1999)
- once irradiation is complete, soils should be handled aseptically until testing

- if soils cannot be tested immediately they should be stored at -20°C or -80 °C until use to suppress re-growth of senesced organisms that survived the irradiation

For any sterilization method, a sub-sample of treated (sterilized) soil should be used to confirm that the microbial population is sufficiently reduced for the purpose of the test design or DQO. This may be done by CLPP (see Table 2), heterotrophic plate count or by the most probable number (MPN) method (Germida, 1993). The test for microbial population in the soil can be performed on a sub-sample of soil before and after the sterilization method is applied. A predetermined acceptable difference (determined by researcher based on study objectives) should be checked to verify if the sterilization was successful.

The reduction of indigenous microbial populations in soil requires significant effort and can change intrinsic soil characteristics, thereby significantly disturbing the integrity of soil samples. Autoclaving soil can: destroy the soil structure, release ammonium-N and amino acids (Trevors, 1996), increase the soluble fractions of organic matter, carbohydrates, NH_4^+, Ca^{2+}, and Mg^{2+}; increase the percent sand and decrease the percent clay fraction of the soil (because of increased aggregation of clay particles); increase the surface area of soil particles; and, slightly increase soil pH and cation exchange capacity (Lotrario *et al.*, 1995). Irradiating soils can increase the availability of nitrogen, phosphorus, sulphur and can release manganese, ammonium, soluble carbon, organic nitrogen, and phosphorus from soil and soil organic matter. In addition, the high energy of the radiation induces chemical reactions in the soil such as the depolymerization of cellulose and the creation of free hydrogen and hydroxyl radials that are strong reducers and oxidizers and cleave C-C bonds (Trevors, 1996).

5.6.3 Dry Heat Sterilization

In addition to autoclaving, sterilization of soil can be achieved by heating soil samples to 200°C for a minimum of 24 hours. To be effective, as with autoclaving, it is important that soils be subjected to dry heat sterilization as thin (e.g., 1 cm) layers. It should be noted, however, that bacterial spores are more resistant to dry heat sterilization than autoclaving and samples might need to be hydrated and incubated for 1 to 2 days prior to dry heat sterilization.

5.6.4 Microwaving Soils

Microwave treatment can be a quick (1 to 5 min) way to reduce, but not eliminate, nematode and fungal populations in soil. Microwave treatment is rarely, if ever, used to sterilize soils; however, if used, it is more effective when soils are sterilized and when moist rather than dry (Trevors, 1996).

5.6.5 Chemical Sterilization

The use of chemicals, which include volatile (e.g., ethylene oxide, chloroform, propylene oxide, methyl bromide) and non-volatile (e.g., mercuric chloride, sodium or potassium azide, formaldehyde, antibiotics like triclosan) compounds to reduce microbial populations in soil (Lotrario *et al.*, 1995, Trevors, 1996) is not recommended as this practice introduces a confounding factor into biological tests. It is a confounding factor because the sterilants themselves are soil contaminants and also because there is the potential for the steriliants to interact with and modify the bioavailability and toxicity of the contaminants of concern.

Regardless of the method of sterilization used, a sterility check of the soil is necessary to ensure that the sterilization process was successful. Further information on methods for checking the sterility of soil can be found in Lotrario *et al.* (1995) and Trevors (1996).

Regardless of the method used to reduce the indigenous microbial population, following treatment ("sterilization"), a subsample of soil should be evaluated to confirm that the microbial population is sufficiently reduced to meet the purposes of the study objectives or DQOs. This may be done by the CLPP (Table 2), heterotrophic plate count, or the most probable number (MPN) method (Germida, 1993). Tests to evaluate microbial populations in soil can be performed on a subsample of soil before and after the "sterilization" method is applied. A pre-determined study-specific acceptable difference of microbial populations before and after sterilization should be used to

determine if the method used to reduce the indigenous microbial population in the sample was effective.

5.6.6 Reducing Indigenous Meso- and Macrofauna

In order to conform to the study objectives and meet the DQOs it is sometimes necessary to reduce the indigenous macroflora and fauna in soil (e.g., seedlings, earthworms, insect larvae). This is particularly true when conducting microcosm tests with introduced test organisms. Indigenous macroflora and fauna can be reduced or eliminated by freezing (-20°C), drying, or microwaving the soil samples (Subsection 5.6.4) (Spurgeon *et al.*, 2002).

Freezing and thawing cycles or air-drying can also be used to reduce the macrofauna and macroflora in intact soil cores. However, freezing and thawing intact soil cores can disrupt the structure of the samples.

One or more subsamples of soil should be collected for analyses of the contaminant(s) of concern before and after manipulation to measure loss of or changes to contaminants through manipulation and to measure the exposure concentrations at the beginning of the test(s).

Section 6

Special Considerations on the Collection, Handling, and Preparation of Soil from Canada's Ecozones

This section provides guidance for sampling, handling, transporting, storing, and preparing soil from ecozones in Canada for which *some* of the universal collection, handling, transport, storage, and preparation procedures provided in Section 3 are not applicable or appropriate. All of the guidance described in previous sections of this document *does* apply to these soils *except* for the circumstances and soils indicated in this section.

Canada is comprised of a wide range of terrestrial ecozones including temperate rainforests, grassland prairies, temperate deciduous forests, boreal coniferous forests, taiga, and arctic tundra. There are 15 terrestrial ecozones in Canada in total, each with a well-defined geographic distribution (Figure 20).

Currently available national and international standardized soil toxicity test methods have been developed for the assessment of soils with neutral to near-neutral soil pH (e.g., 6 to 8) and organic matter content ranging from approximately 3% to 12%. This description generally characterizes Ah horizons of agricultural soils in Canada (prairies and mixed-wood plains ecozones) and deciduous mixed forest ecoregions soils in the southeastern part of the country (mixed-wood plains) (EC, 2003a). These methods might also be applicable to some soils distributed within the montane cordillera and Atlantic and Pacific maritime ecozones as well. The universal procedures described in Section 3 also apply to these types of soils. However, these soils represent a relatively small portion of Canadian terrestrial ecosystems.

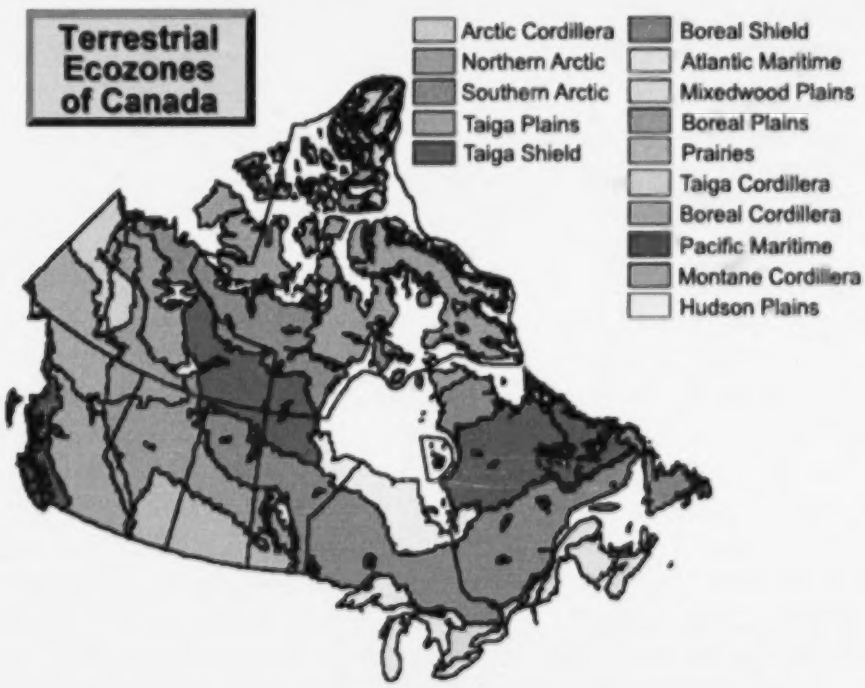


Figure 20. The terrestrial ecozones of Canada (reproduced from EC, 2005b).

There are many other soils in Canada with widespread distributions that have properties that fall outside of the ranges considered typical by standard methods and that also require some special sampling, handling, transport, storage, and preparation procedures.

The distribution of some of the 10 Soil Orders of Canada (Subsection E.2; Appendix E) roughly corresponds to that of one or two ecozones (e.g., the distribution of the Chernozemic Soil Order corresponds closely to that of the prairies ecozone); however, in most cases the Soil Orders are either distributed in small areas across the country or their distribution extends beyond two or more ecozones. More information on the geographical and ecological distribution of each of these Soil Orders can be found in Appendix E. Table E.1 (Appendix E) provides descriptions of the 10 Soil Orders, their diagnostic horizons, and where they predominantly occur; Figure E.4 (Appendix E) provides photographs of soil profiles for each of these Orders and Figure E.5 (Appendix E) provides maps of the geographical distribution of each of the Soil Orders in Canada.

The most widespread soils in Canada for which currently available test methods, and some of the universal collection, transport, storage, and preparation procedures described in Section 3, do not apply include boreal forest soils, stony/ shallow soils, organic soils, cryosolic soils, and wetland soils.

Forest soils, such as luvisolic, brunisolic, and podzolic soils found in the boreal shield, boreal plains, boreal cordillera, taiga shield and taiga plains ecozones (Figure 20; Table E.1, Figures E.4 and E.5; Appendix E) are highly stratified and are often overlain by a highly organic LFH or O horizon. Some of these forest soils have very low pH (< 4) such as podzolic soils underlying much of the boreal ecozones. Throughout all of Canada's ecozones there are areas ranging from small pockets to large regions where soils can be very shallow (only a thin layer of soil over the bedrock) or stony due to glacial deposition. Organic soils are predominantly found in boreal forests and wetlands in the Hudson and boreal plains, taiga and boreal shield and sporadically elsewhere in Canada (Figure 20; Table E.1, Figures E.4 and E.5; Appendix E). Organic soils are comprised primarily of organic materials in varied states of decomposition and are commonly known as peat, muck, bog or fen soils. Sampling and handling of organic soils (Subsection 6.4) is restricted to discussions of terrestrial organic soils, whereas wetland soils (e.g., peat, bog, fen, marsh, swamps) are discussed in a separate subsection (Subsection 6.5). Wetland soils cover a significant portion of Canada's landmass, and are distributed throughout every ecoregion of the country (NRCan, 2009b). The majority of wetlands in most regions are peatlands, which are found in the southern arctic, taiga plains, taiga shield, Hudson plains, and boreal cordillera ecozones. Despite the fact that wetlands are transitional ecosystems comprising of both semi-aquatic and semi-terrestrial characteristics, the widespread distribution of wetland soil in Canada and the concomitant potential to be contaminated as a result of anthropogenic activities warrants a brief discussion herein. Cryosolic soils are arctic, subarctic forest, and tundra soils that cover most of the northern third of Canada where *permafrost* exists close to the surface of both the mineral and organic soils. Cryosolic soils predominate in the northern and southern arctic and arctic cordillera ecozones, are common in taiga plains, shield and cordillera and Hudson plains ecozones (Figure 20; Table E.1, Figures E.4 and E.5; Appendix E), and extend into the boreal forest and alpine regions (AAFC, 1998).

Given that boreal forest, cryosolic, shallow or stony, organic, and wetland soils cover most of Canada's land mass, and that significant past and current anthropogenic activities in these ecozones (e.g., mining, forestry, oil and gas production) have created, or have the potential to create, contaminated lands, it is important that these soils are properly collected, handled, and prepared for biological testing.

6.1 Boreal Forest Soils

Most of the guidance provided in Section 3 is applicable to the collection, handling, transport, storage, and preparation of boreal forest soils. However, there are some specific modifications to some sampling, storage, and preparation procedures and some special considerations for test species selection when testing boreal forest soils.

6.1.1 Sampling Boreal Forest Soils

Many forest soils, including boreal, mixed wood, and deciduous, have thick organic horizons at the surface and are distinct from organic soils on the basis of the thickness and organic carbon content of their organic (O) horizon or of their luvic, fulvic, and/or humic (LF and/or H) horizons (AAFC, 1998). Often this surface layer is of interest to the contaminated site assessor; either because of the concentration of contaminants in the top few centimetres of the forest floor and/or because of the ecological importance of these highly organic and biologically *active layers*. Evaluation of contaminant-related changes in the upper organic horizon is often a very important component of studies of contaminant loading via airborne transport pathways, accompanied by wet and/or dry deposition. Other contaminant release mechanisms that result in loadings to organic surface horizons of soils include spills/releases in winter onto the surface of snow and ice which subsequently melts and infiltrates in warmer weather. Whether or not surface organic layers are sampled depends on the objectives and the DQOs of the study.

Sampling the thick layer of organic material at the surface of the soil requires specific sampling devices (Table G.1), and procedures slightly different from those described in Subsection 3.6.5 for mineral horizons. Recommended procedures are as follows:

1. Establish the boundaries of the sample location as per guidance in Subsection 3.3.8.

2. Clear the soil surface of loose materials (debris) by hand or by gentle raking (Figure 21).

3. If extensive vegetation covers the surface of the plot then cut the vegetation to the surface of the soil where the sample is to be collected. In many areas, there may be an appreciable layer of actively growing or recently killed sphagnum or brown moss overlying the soil. Soil samples should exclude the active growing as opposed to detrital bryophyte (*Sphagnum* or moss) material, although there is often little to distinguish this transition (step 4v).

4. If the organic material is extracted as a large (e.g., > 1 L) sample volume (adapted from Bélanger and Van Rees, 2008):

    i. Place a sampling frame or a reference frame (which can range from 100 to 900 cm^2) onto the soil surface.

    ii. Push the frame into the forest floor until the mineral soil is reached. If the frame has a sharp cutting edge on the bottom surface, it can be driven into the organic layer using a mallet or a nylon-headed hammer.

    iii. If the L (litter) horizon is to be sampled separately from the other organic (FH) horizons (depending on the study objectives), the litter should be collected at this step from within the sampling frame.

    iv. Cut through the forest floor material using a corrugated knife or shears on the inside edge of the frame.

    v. Once the sample is cut on all sides, partition the organic layer from the mineral soil using a stainless steel spatula, forceps, scoop, shovel, or a trowel.

- If the forest floor material is very cohesive and easily separated from the underlying mineral A horizon, cut the layer away with a knife or trowel and simply roll the organic material up like a carpet.

- If it is difficult to distinguish between the lowest (H) organic horizon and the top (Ah) mineral horizon, bias might be introduced into the samples as some Ah material can be incorporated into the organic layer samples. If it is critical to prevent inadvertent inclusion of Ah material in organic layer samples, pre-sampling analyses of the carbon content of the different soil layers is recommended (Bélanger and Van Rees, 2008). If site access is difficult, the layers can be separated by colour and texture.[34]

vi. Place extracted soil into a sample container(s). If soil is to be prepared on-site (dried, homogenized, or sieved before shipping the sample to the laboratory), place the soil sampled onto a plastic sheet, cotton sheet (if contaminant(s) are plastic related), tarp or a receiving container(s) until the entire sample is collected and ready for preparation.

5. Typically surface horizons in forest soils will contain numerous roots; if this is the case, once the sample is collected gently tap the soil from the roots into the sample container and either discard the plant material, keep it with the soil sample, or collect it as a separate sample depending on the study objectives (USEPA, 2006).

6. Sample the underlying mineral horizons (using procedures described in the universal procedures Subsection 3.6.5) using shovels or trowels or other sampling devices appropriate for sampling mineral horizons as follows:

   i. Extract the soil from the top mineral horizon (the A horizon) in scoops using trowels or shovels to the depth of colour change.

   ii. Place extracted soil into a sample container(s). If soil is to be prepared on-site (e.g., dried, homogenized, or sieved before shipping the sample to the laboratory), place the soil sampled onto a plastic sheet, cotton sheet [if contaminant(s) are plastic-related], tarp, or a receiving container(s) until the entire sample is collected and ready for preparation.

   iii. Ensure that all the soil from the horizon or depth is removed from the entire plot.

   iv. Repeat steps 6 (i) and 4 (iii) for underlying B and C horizons if they are to be sampled (based on study objectives).

7. If smaller sample volumes are required, or if consolidated samples (e.g., intact cores) are required, extract the sample using corers or other soil sampling devices (Table G.1) that are designed to collect organic soil samples (these usually have wide diameters and a sharp cutting edge to slice through fibrous organic material). Extract the soil as follows (adapted from Mason, 1992 and ISO, 2003a):

   i. Corers with wide diameters (e.g., 9 to 30 cm) and that have a sharp, serrated cutting head are recommended. Sharp serrated cutting heads on corers are useful as hand-pushing with a rotating motion helps cleanly cut through smaller roots and fibrous materials.

[34] In these cases, Bélanger and Van Rees (2008) comment that identification of humus layers can be complex and refer their readers to Klinka *et al.*, (1981) and Green *et al.* (1993).

ii. Drive the corer (using a mallet or steel hammer with a nylon head) to the known depth of the organic layer horizon or to the texture or colour change that indicates the horizon transition (a "test" core might have to be extracted from the plot first to determine where the texture or colour change occurs in very heterogeneous sites) or to the desired depth.

iii. Avoid tilting the corer or ring while pushing it into the forest floor.

iv. Extract the sample. Various approaches can be employed to limit the loss of the cored soil during extraction of the corer from the substrate. One set of techniques involves creation of suction in the top of the corer barrel with a plunger or larger rubber stopper. At times, it may be necessary to hand excavate around the outside of the core barrel, allowing sufficient access to secure the bottom of the corer barrel prior to withdrawal.

v. Cut away any transitional soil or roots from the bottom of the sample

vi. For unconsolidated samples, extract the sample from the corer and place the sample into the sample container.

vii. For consolidated samples (intact cores), cap the core liner and place it into a container for storage and shipment.

viii. For both unconsolidated and consolidated core samples, the depth of material retained should be recorded along with the depth of penetration of the core barrel. This allows for an estimate of soil compression or loss during coring, and provides information essential to subsequently relating biological test results to the original depth and thickness of the strata sampled.

ix. If the sample is to be prepared on-site (e.g., dried, homogenized, or sieved before shipping to the laboratory), place the extracted sample onto a plastic sheet, cotton sheet [if contaminant(s) are plastic related], tarp or a receiving container(s) until the entire sample is collected and ready for preparation.

x. Repeat steps 7 (ii) to 7 (ix) until all the soil required is extracted from the plot and placed in the same sample/receiving container.

xi. To extract soil from mineral horizons (A, B, or C), repeat steps 7(ii) to 7 (x) using sampling devices appropriate for mineral horizons.

8. Boreal forest soils are often relatively shallow compared with other soils and are much less likely to be collected at depth. However, if soil samples are collected at depth, extracting samples using a drill truck (e.g., solid-stem auger, split-spoon push rig, sonic rig) can be a more efficient and less labour-intensive method. Access to the sampling locations through a forested site, however, can be a challenge. In addition, some drilling apparatus and techniques have a large potential to modify the soil matrix and introduce artefacts that can influence biological test results. Air rotary drills, solid-stem auger techniques, and those that use various drilling fluids should be avoided. However, standardized methods published by the American Society for Testing and Materials (ASTM) exist and might be appropriate depending on the study objectives and/or DQOs. Interested readers are directed to the following references: ASTM (2008a,b,c; 2009a,b).

Normally very large sample volumes (e.g., 100s of litres) of forest soils are not collected as this represents significant forest habitat

degradation which would be at odds with the usual objective of contaminated site assessments and/or remediation, which is (in part) ecosystem protection. Indeed, a common reason for conducting biological testing of contaminated forest soils, rather than excavating all contaminated soil, is to determine the relative risk to ecological receptors of leaving contaminated soil *in situ* compared with excavating contaminated soil, as the latter would result in significant forest habitat destruction.

Guidance on preparing soils on-site is similar to that in the universal procedures and the reader should refer to guidance in Subsection 3.6.6.

Figure 21. Sampling location within a boreal forest with loose surface vegetation removed (photo: B. Smith).

6.1.2 Forest Soil Variability

Forest soils in general tend to have greater variability in their physical and chemical characteristics than many other soils except cryosols. This variability is due to bedrock type, parent material, tree species composition, understorey vegetation composition, disturbances (e.g., fire, tree fall), and forest management activities. Boreal forest soils developed from the shallow rocky till of the Canadian shield are very heterogeneous; soil horizon thickness varies as a continuum with complex spatial patterns (e.g., thick FH material in pits and A horizon pockets along old root channels) (Figure 22) (Bélanger and Van Rees, 2008). Many of these factors can also result in greater variability in soil contaminant concentrations and contaminant toxicity. Guidance specific to issues encountered due to the shallow nature of some boreal forest soils is provided in Subsection 6.2.

Studies have demonstrated that the results of bulking samples from individual sample locations provide good estimates of the real soil contaminant concentrations and soil properties in forest soils. A general rule is that sampling a larger surface area will reduce the microsite variability in a forest sample (once it is dried, sieved, and homogenized) and a minimum sample location area of at least 200 cm^2 is recommended (Bélanger and Van Rees, 2008). However, scale is an important issue in light of the complex spatial variability in boreal forest soils, and needs to be considered in the context of the ecological receptors of interest. On the one hand, soil conditions over areas with a diameter much greater than one metre would best capture possible effects on large woody shrubs and trees such as black spruce or paper birch. On the other hand, the natural variation in characteristics of soil mesofauna over distances of even a few centimetres laterally or vertically is likely to be very large, and challenging to accommodate in a contaminant assessment program.

6.1.3 Storage

The LFH and A horizons of boreal forest soils typically have high organic matter content, relatively high field moisture content, and high microbial populations. As a result, it is recommended that these soils be prepared for testing (e.g., air-dried, sieved, and homogenized) before storage and are stored at cool temperatures (e.g., 4°C) until testing (Subsections 3.9.2 and 3.10). These actions will retard the proliferation of unwanted microbial growth such as fungi that might influence test species performance; however, any preparations should be conducted in accordance with study objectives and/or DQOs (e.g., if microbial testing is to be conducted the preparations suggested here are not recommended).

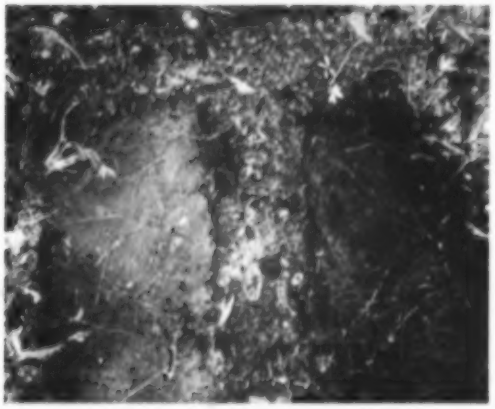


Figure 22. Two adjacent soil plots both excavated to 10 cm at a forested site with very shallow soils within the boreal shield ecozone (photo: G. Stephenson).

6.1.4 Preparation

As already mentioned, the LFH horizons of boreal forest soils typically have very high organic matter content and can contain a large quantity of plant roots. If soil contamination is associated with the LFH horizon then it is prudent to test soils with all of the organic material present, and the soil reconstitution procedures described in Subsection 3.10.3.6 should be followed if necessary.

6.1.5 Test Species Selection

Because many of the procedures for sampling, handling, storing, and preparing soil collected from boreal forest ecosystems are similar to the universal procedures, in some cases special consideration might not be required when conducting biological tests with boreal forest soils. However, in many instances, the soil pH is very low (pH < 4), especially for soils that developed on the Canadian Shield (e.g., within the boreal shield ecozone), the soil pH is very low (pH < 4). Most test species recommended by standardized toxicity tests, including those in the Environment Canada terrestrial test methods (EC, 2004a, 2005a, 2007a) are intolerant or perform sub-optimally in soils at this low pH such as enchytraeids (ISO, 2004a) (Jänsch *et al.*, 2005). Adjusting the soil pH of these soils to accommodate the needs of currently standardized test species is not appropriate and will introduce artefacts into the toxicity test data. Instead, the recommended action is to select test species that are naturally tolerant of low soil pH. Environment Canada is currently developing terrestrial toxicity test methods for boreal forest soils with test species adapted to boreal forest conditions (Moody, 2004, 2006; Römbke *et al.*, 2006b). These species include earthworm, arthropod, and woody and herbaceous plant species (EC and SRC, 2007).

6.2 Shallow/Stony Soils

Shallow rocky soils are common throughout Canada, and very shallow, rocky, and/or very stony soils are particularly common in the boreal shield, taiga, and arctic ecozones. These soils can be difficult to sample and special consideration is warranted when developing sampling plans or sampling sites with these soils. These types of soils can be further described as follows (AAFC, 2007):

- ***Rocky soils*** occur where there are rock out-croppings and/or only a shallow layer of undeveloped soil covering the bedrock (this is also referred to as shallow soil). Rocky or shallow soils are commonly found throughout the boreal and taiga regions of Canada. Shallow soils are also found in alvars, located mostly in the Great Lakes basin of the mixed wood plains ecozone. Alvars are areas of flat glaciated limestone bedrock; soils are thin or absent and these ecosystems usually support unique plant species and species associations (Nature Conservancy Canada, 2009).

- ***Fragmented soil*** or soils containing coarse fragments are soils in which stones or rocks are found imbedded throughout the soil column (not just at the surface but also at depth). In Canada, coarsely fragmented soil is usually of glacial origin such as glacial till deposits or of glaciofluvial origin, located in outwash plains and deposits. Glacial till deposits can be quite irregular or can be accumulated into a ridge (glacial moraines). Glacial deposits can be further classified by the manner in which they

were transported and deposited (NRCan, 2009a).

- ***Stony soils*** are soils where stones or rocks are found on or in the surface layer of the soil. This is usually due to glacial deposits or in areas of past fluvial deposits and re-working, where historical riverine action and meandering channels has resulted to varying degrees in sorting of rock and sediment by size. Stony soils can be found regionally throughout all of the ecozones in Canada (NRCan, 2009a).

6.2.1 Sampling Shallow/Stony Soils

Rocky and coarsely fragmented soils can pose a particular challenge for site assessments because the biotic contaminant exposure potential is unclear. This is because the potential for contaminant exposures can theoretically occur at some (poorly predictable) point along a spectrum bounded by the bulk soil concentration at one end and the concentration of contaminant in only the finest soil fractions present at the other end of the spectrum. In addition, normal sampling tools may not be able to penetrate the soils (tools may hit rock) or there may not be sufficient quantities of soil for sampling requirements. Insufficient soil may be a result of shallow soil or a result of stony soils (reduced quantity of soil upon removal of stones).

In most cases if the rocks or stones are only on the surface (stony soils) they can be removed prior to sampling. There are circumstances, however, when removing stones is not appropriate, for example, if the source of contamination is airborne and removing surface material (including rocks and stones) will bias the concentration of contaminant(s) in the sample. Bias of the contaminant concentration may also be a concern when sampling coarsely fragmented soil, if removing the fragments will significantly alter the sample volume or produce a sample that is not representative of the site (Mason, 1992). Removal of fragments from the soil is also not appropriate if the source of contaminant makes up part of the fragments found in the soil (e.g., explosive residue associated with former range, munitions facilities or military bases, or larger lead particles in soils at firing ranges) as removal of the fragments may in fact remove the contamination. The source of contamination, study objectives, and the DQOs should be considered at the study planning stage in order to facilitate well-informed decisions about removing stones or rocks from samples.

There are three recommended approaches to sampling rocky/shallow, stony, and fragmented soils:

- Relocation to a different sample location. In some cases there may be an adjacent pocket of deeper soil (in the case of rocky/shallow soil) or a less stony, or fragmented soil (due to the sporadic nature of glacial deposits). DQOs must be considered when sampling these deeper pockets to ensure representative soil samples are collected. The soil survey should help reduce the risk of encountering a particularly shallow or fragmented soil, and also help determine if there are suitable alternate sampling sites (Subsection 3.3.3.2). It may not be possible find an alternate sampling site (e.g., if the extent of rocky/fragmented/stony soil is greater than the extent of contamination). If this is the case, the objectives and DQOs of the study must be re-evaluated in consultation with all parties involved (e.g., project managers, regulators, stakeholders).

- Collect larger volume bulk samples and sieve to obtain sufficient mass of the soil particle size fraction of interest for biological testing. This can be done either by excavating a pit that is larger than the largest rock in the sample or by collecting smaller-sized samples over a large area using a corer (Belanger and Van Rees, 2008). Care must be taken not to strip all of the soil from the site.

- Conduct biological testing of the entire bulk sample, with gravel and small boulders. This would not be a valid approach for most available test species and methods, and should only be carried out with biological test species that would normally interact with coarser materials in

the soil over larger areas (e.g., larger tree seedlings with root and rootlet penetration throughout the stony matrix). A particular challenge with biological testing of large soil samples with appreciable gravel and small boulder content is that finer soils tend to become vertically re-distributed during transport, storage, and preparation.

Some specialized tools are used for sampling coarsely fragmented soils. These include augers, which are reinforced to be able to penetrate through rock fragments, and corers designed with a larger circumference so that when coring, rock fragments are included in the core, thus allowing the core to penetrate to greater depths (Page-Dumroese and Brown, 1999). Appendix G provides descriptions of various sampling tools.

As with any site that is to be sampled, care must be taken to minimize disturbance when sampling sites with shallow soils. This is especially true for sites that represent unique habitat and/or provide habitat for vulnerable species. Depending on the nature and sensitivity of the site, the presence of vulnerable habitat and/or species, the objectives of the site assessment, and, applicable regulations, collection of soils in sensitive habitat may not be recommended or might in fact be prohibited. This information should be obtained when collecting background data during the development of the study plan and be corroborated with the site survey. If it is determined that the site to be sampled is a sensitive site, management decisions should be made *a priori* in consultation with all stakeholders (including regulatory personnel).

6.2.2 On-site Handling, Storage, and Transport

On-site sample handling, storage, and transport guidance provided in Subsections 3.6.6, 3.7, 3.8 and 3.9 are applicable for rocky, fragmented, and stony soils. Note, however, that for stony soils containing minimal fine-textured particles there is a greater tendency of the finer material to settle out during transport and storage and/or be lost during preparation.

6.2.3 Preparation

As discussed in Subsection 3.10.3.3, careful consideration of the size of particles to remove during soil preparation is necessary. This is particularly true for stony, fragmented and/or rocky soils. Biological test methods tend to focus on soils in the sand-silt-clay size fractions (< 2 mm effective diameter based on most classification schemes), as opposed to gravels and boulders. There may also be cases where a smaller particle diameter cut-off is appropriate for defining the fraction of stony soils of relevance to various ecological receptors [e.g., < 1 mm; < 250 µm (silt-clay fraction)]. Regardless, it is important that the sample collection and preparation methods adequately address how contaminants are likely to be distributed across soil particle sizes within the larger soil matrix, and how various biota of interest would interact with the stony soil in its "natural" state (e.g., in the field). If finer particles contained within the interstices of larger gravel and boulders comprise the major available growth medium for the soil, then it is important that this fraction be the focus of testing. Depending on the type and source of the contamination, removal of rocks or stones could alter the concentration and/or characteristics (e.g., bioavailability) of the contamination in the soil. For example, Vischetti *et al.* (2010) found that organic carbon found in rock fragments > 2 mm contributed to pesticide adsorption and degradation in the environment. Therefore removal of these fragments for testing would change the bioavailability of the pesticides in the soil, potentially effecting biological testing results.

If microbial testing is to be conducted, removal of stones or rocks should be considered with regard to study objectives and/or DQOs. The removal of rock fragments from field-collected soils can result in the underestimation of the total amount of microbial biomass and a misrepresentation of the structure of the microbial community (Certini *et al.*, 2004). Removal of gravel and stones can also profoundly influence water-filled and vapour-filled porosity, hydraulic conductivity, moisture retention potential, and

degree of oxygen penetration of soil. This can all influence the rates of diffusion of a large suite of natural substances and contaminants transported in soil gas or interstitial water.

6.3 *Cryosolic Soils*

Most soils are frozen during some part of the winter in Canada; soils collected for ecotoxicity assessments soils are usually (but not always) sampled before the surface soil freezes. However, there is a group of soils that covers most of the northern third of Canada where permafrost exists close to the surface of both the mineral and organic soils. These soils are called cryosols and they predominate in the tundra, are common in the taiga, and extend into the boreal forest and alpine regions (AAFC, 1998). These soils often undergo cryoturbation, meaning that cyclical freezing and thawing results in disruption of the soil horizons causing patterned ground features such as circles, polygons, and earth hummocks (Table E.1, Figure E.2; Appendix E) (AAFC, 1998; Tarnocai, 2009). Cryosols are unique soils with a wide variety of characteristics; they can be organic or mineral soils, very acidic or neutral, very clayey, or very sandy. Because of their unique characteristics, cryosols require special sampling designs, and when frozen, sampling procedures and sampling tools.

6.3.1 *Sampling Cryosols*

Cryosols have an active layer, which is the upper layer that annually freezes, and thaws, and the lower perennially frozen layer (permafrost). The diagnostic characteristic of cryosols is that the permafrost lies within 1 to 2 m of the surface of the soil (AAFC, 1998). Because of cryoturbation, the variability of soil properties in cryosolic soils is typically very high. This will, in turn, increase the heterogeneity of the horizontal and vertical distribution of contaminants (since freeze-thaw cycles induce massive fracturing, contaminants can be physically relocated and/or move through preferential pathways that result from the fracturing) as well as contaminant bioavailability and toxicity.

Recent investigations into the implications of this variability on biological test results suggest that when sampling cryosolic soils, significantly greater sampling effort (e.g., > 30 samples per site) might be required in order to detect small differences in test organism performance (Anaka *et al.*, 2008). However, guidance for the use of sampling strategies and determining the number of samples to collect for contaminated sites in tundra or taiga ecosystems, or any contaminated site in Canada containing cryosols, is still in early stages of development.

The actual collection, handling, transport, and preparation of samples from the active layer of mineral cryosolic soils during the times of year when it is not frozen is similar to that of other mineral soil types (Anaka *et al.*, 2008). Therefore, other than Subsections 3.3.4 (Selection of Sampling Locations) and 3.3.9 (Sample Size), the guidance provided under the universal procedures in Section 3 and/or that for boreal forest soils (Subsection 6.1) and terrestrial organic soils (when mineral cryosols have an organic surface layer) (Subsection 6.4) is applicable.

6.3.2 *Sampling Frozen Active Layers, Permafrost Layers, or Frozen Non-Cryosolic Soils*

Although it is unlikely that depths greater than 60 cm are of interest for the purposes of biological testing, it is conceivable that there are times when sampling the permafrost layers of cryosols is desired. It is also conceivable that for specific study objectives or constraints (e.g., limited access to the site in warmer seasons); frozen active layers of cryosolic soils or frozen non-cryosolic soils need to be sampled.

Collecting frozen consolidated or unconsolidated samples is quite similar in some respects to sampling non-frozen soils; only the tools differ. Tools used for sampling frozen soil are stronger, more durable, and usually have robust, sharp serrated edges to dig through hard frozen soil, peat and ice.

Some of these tools are listed in Table G.1 (Appendix G).

If frozen bulk samples are collected from pits or trenches from mineral soils, then a pickaxe is used to remove blocks of frozen material from the sample location. A shovel can also be used to collect the smaller soil chunks loosened by the pickaxe. In addition to a pickaxe, electric hammers can also be used to create a trench or pit in the soil (Tarnocai, 1993) from which samples can be collected. If the bulk samples are collected from thick organic layers (frozen peat), then an efficient method is to use an electric chain saw, which can cut blocks in a grid pattern for sampling. Another device is then used (e.g., an electric hammer) to cut away the base of the peat blocks to release the sample from the ground. An electric chain saw should be used rather than a gasoline-powered saw to prevent contamination of the samples, or buildup of toxic fumes if the chain saw is operated in a trench. It should be kept in mind that oil used for chainsaws (both electric and gasoline-powered) has the potential to contaminate the sample (Tarnocai, 1993).

When sampling the permafrost layer of cryosolic soils, often a pit must be dug first to remove the active layer of the soil (the active layer might or might not be collected as a sample depending on the study objectives). Once the active layer has been removed, specialized augers or corers are applied to the surface of the frost layer to collect samples. When sampling permafrost in relatively warm weather (e.g., between -1°C and 0°C), soil cores with large amounts of unfrozen water can begin to thaw during coring; if these cores also contain high amounts of clay, they can be very difficult or impossible to sample (Tarnocai, 1993).

There are other unique challenges to sampling frozen soils. Often a small depression in the surface of the frozen soil layer needs to be chipped away to provide footing for the soil corer or auger. Depending on the tool, care must be taken to not stop rotating (or drilling) the corer or auger, otherwise it can become frozen to the side of the hole. It is also not unusual for the bottom of the soil core to remain firmly attached to the ground when the coring device is removed; in these instances, the core is removed with a core catcher. The core catcher is attached to an extension and is lowered over the core; quick sideways or twisting actions break off the core and the sample is withdrawn from the soil (Tarnocai, 1993). Once samples have been taken, they can be temporarily maintained at cool temperatures in or on the surface of permafrost.

The universal transport, storage, and preparation procedures described in Section 3 apply to frozen or cryosolic soils, although it is very important that the sample containers are tightly sealed so they will not leak when they begin to thaw. These samples should also be transported to the laboratory as quickly as possible and immediately dried upon arrival so that anaerobic conditions do not develop. If the study objectives dictate that the soils should remain frozen until tested, transport and storage arrangements should be made ahead of time to accommodate this.

6.3.3 Test Species Selection

The selection of test species for biological tests with cryosolic soils must be carefully considered. The physical and chemical characteristics of cryosols can range widely, and some characteristics, like soil pH, might extend to ranges beyond those tolerated by most species recommended by standardized toxicity tests, including those in the Environment Canada terrestrial test methods (EC, 2004a, 2005a, 2007a) (e.g., low soil pH). Therefore, it is prudent to have a good understanding of the characteristics of the soils at the site early in the site assessment process to allow the selection of the most suitable test species.

Most of the test species currently recommended in standardized tests, although potentially able to tolerate cryosolic soil conditions, do not occur in ecosystems containing these soils and therefore lack ecological relevance. Although this is an

unavoidable situation encountered when assessing contaminated sites in northern Canada at present, Environment Canada is currently developing terrestrial toxicity test methods with test species adapted to tundra and taiga conditions (Princz *et al.*, 2010).

6.4 Organic Soils

Organic soils are comprised primarily of organic materials and are commonly known as peat, muck, bog, or fen soils. These soils contain more than 30% organic matter (by weight), are commonly saturated with water for prolonged periods, occur widely in poorly drained depressions, and are derived from the vegetation that grows on those sites (AAFC, 1998). Three of the four types of organic soils in Canada are commonly saturated throughout the year and consist largely of undecomposed organic material. These three soil types (fibrisols, mesisols, and humisols) are products of wetland development. The fourth type of organic soil is the terrestrial folisols, which, as products of upland ecosystem development such as forests, are only briefly saturated with water. Shallow folisols (10-cm depth) occur throughout Canada (AAFC, 1998). Although some corers and probes used might be specially designed to cut through organic material (Sheppard *et al.*, 1993)(Table G.1), many of the methods and procedures used to collect folisols (which occur in forest ecosystems) are similar to those described for sampling boreal forest soils and guidance provided in Subsection 6.1 should be followed.

In addition to terrestrial organic soils (folisols), terrestrial mineral soils can also have organic (O) horizon surface layers that overlay the A mineral horizons. These types of soils can be found in areas adjacent to wetlands; for example, mineral soils overlain with a thick (e.g., 20 cm) peaty organic surface material collected adjacent to a bog. Depending on how fibrous and cohesive the O layer(s) is, procedures for sampling the O horizon of mineral soil might be similar to the universal procedures (Subsection 3.6.5.1) for sampling the A horizon (e.g., extracting the O horizon using shovels or trowels) or similar to those described for sampling boreal forest soils (Subsection 6.1) (e.g., using a cutting frame and shears to remove the O horizon).

Organic soils have many of the same features as the soils found in the boreal forest ecosystem, in particular, low pH. As such, the guidance found in Subsection 6.1.4 for test species selection is also applicable to organic soils.

Wetlands and peatlands, comprised mostly of organic soils, occur in semi-terrestrial to aquatic environments where the water table is seasonally near or above the surface with minimal water flow (UMA, 2008). As a result, the chemical, hydrological, and ecological processes in these soils are complex and different from that of terrestrial organic soils. Guidance specific to sampling, handling, storing, and preparing wetland soils is discussed in a separate section (Subsection 6.5).

6.5 Wetland Soils

Wetlands are transitional ecosystems with both semi-aquatic and semi-terrestrial characteristics. They comprise a significant (18%) portion of Canada's landmass and are dispersed throughout every ecozone of the country (NRCan, 2009b). Compared with terrestrial systems, wetlands experience seasonal water level fluctuations, rapid decomposition rates, and are often comprised of significantly different vegetative communities. These and other factors can result in different fate, behaviour, and toxicity of contaminants in soils in wetland ecosystems compared with those in terrestrial soil systems. Wetlands have a large distribution throughout Canada, and as such, can be contaminated by mining, forestry, industry, and agriculture; as a result, wetlands might be subjected to contaminated site assessments. It is important, therefore, that these soils be properly collected, handled, and prepared for biological testing.

Wetlands can be defined as "areas where soils are water-saturated for a sufficient length of time such that excess water and resulting low

soil oxygen levels are principal determinants of vegetation and soil development. Wetlands generally have an abundance of hydrophytes (aquatic plants) in the vegetative community featuring (hydric) characteristics" (Mackenzie and Moran (2004) as cited in UMA, 2008). Wetlands cannot be defined as either exclusively terrestrial or aquatic; as such they are unique and important ecosystems. In Canada, there are five classes of wetlands (Zoltai and Vitt, 1995; UMA, 2008):

- ***Bogs*** receive water only from rain and snow melt and are dominated by *Sphagnum*. Bogs are acidic (often pH < 4) and nutrient-poor, the base cation content is limited, and water flow is restricted. These characteristics primarily result from decomposition processes (acidification from the production of humic acid) and the production of *Sphagnum* (nutrients bound by peat). Bogs can be open, wooded, or forested.

- ***Fens*** are influenced by the chemistry of the surrounding upland mineral soils and are dominated by bryophytes (e.g., mosses). Fens receive water from the surrounding uplands and groundwater. Fens can be sub-categorized according to their pH as rich (pH > 7), moderate-rich (pH 5.5 to 7.0) or poor (pH < 5.5). Rich fens can be similar to swamps (except in hydrology) but are dominated by brown mosses and an abundance of sedges. Likewise, poor fens can be very similar to bogs (except in hydrology) and dominated by *Sphagnum*.

- ***Swamps*** are influenced (like marshes) by groundwater and surface water flows, and are therefore subject to strong seasonal variations in water levels. Swamps are characterized by tall trees or shrubs with little peat accumulation. Little to no peat accumulation occurs in swamps due to high decomposition rates and seasonal lowering of water levels. Where the groundwater is acidic, and prolific growth of trees inhibits water flow swamps can become acidic.

- ***Marshes*** are strongly influenced (like swamps) by groundwater and surface water flows and are subject to large seasonal water level fluctuations. Marshes are open, non-peat forming wetlands dominated by sedges and other monocots. They tend to have high concentrations of phosphorous and nitrogen with a high rate of decomposition of organic matter. Marshes generally have high biodiversity. Marshes can be saline or alkaline and are generally well buffered.

- ***Shallow open water*** is defined by non-peat forming wetlands where the depth of water is less than 2 m of water at mid-summer. Influenced by adjoining aquatic systems and often a transitional area, this wetland type is dominated by submergent floating plants. The water and soil chemical characteristics of shallow open water wetlands are not diagnostic.

In Canada, most of the wetlands are peatlands (bogs and fens); therefore, the soils in these wetlands are usually organic. Natural mineral wetland soils can be silty or clayey; however, artificial wetland soils tend to have more sand (Zoltai, 1978). Many of the wetlands in Canada are found in the southern arctic, taiga plains, taiga shield, Hudson plains, and boreal cordillera ecozones (20% of wetlands in Canada are in the arctic and 5% of arctic land is wetland) (NRCan, 2009b). Permafrost conditions may be encountered in wetlands in northern ecozones, in fact, the low thermal conductivity of bogs might result in the maintenance of permafrost and the extension of discontinuous permafrost through the southern limit of boreal peatlands (UMA, 2008). See Subsection 6.3 for guidance on sampling frozen soils, including permafrost layers. Organic wetland soils such as peat occur extensively in Canada, particularly in peat deposits dominated by sphagnum mosses (AAFC, 1998). These peat deposits are extensively mined, especially in Quebec, for horticultural purposes, although some investigation has occurred into the widespread use of peat as an alternative energy source (Monenco, 1981). Commercial methods have been developed for the large-scale extraction of wet and dry peat for industrial purposes; however, these are generally not applicable for sampling for contaminated land assessment

unless the mined peat land is the contaminated site.

Wetlands can be natural, as just described, or artificial. Artificial wetlands are often constructed for remediation purposes (Campbell *et al.*, 2002). The natural filtration properties of wetlands are re-created so they can act as biofilters for contaminated water or wastewater. In most cases, the sampling of an artificial wetland is relatively straightforward as sample sites (for soil and water) are engineered into the design of the wetland to facilitate monitoring.

6.5.1 Field Measurements and Observations

6.5.1.1 Site Description

A thorough site description is very important if sampling wetlands (Subsections 3.4 and 3.5). Information such as site hydrology, topography, dominant vegetation, season in which sampling occurs, amount and condition of water (standing or flowing), and soil parent material will determine the type of wetland being sampled and help characterize the site. This information, much of which is usually collected early in the site assessment (Subsection 3.3.3), can then be used to understand the physicochemical characteristics of the site and determine the type of biological test and/or test organisms most appropriate to assess contamination at the site.

It is recommended that on-site chemistry measurements be taken, as wetland soil chemistry can change very quickly after being sampled. The water level at the time of sampling will influence soil physicochemical characteristics and microbial community composition. The following measurements are recommended to help determine the type of wetland being sampled and define the undisturbed state of the soil:

- pH,
- dissolved oxygen,
- *redox potential*,
- water level
- weather, and,
- time of year

In addition, sampling iron, other metals, sulphate, and sulphites should be considered as wetlands are known to accumulate metals and sulphur (Wieder and Lang, 1986).

6.5.1.2 Identification of Sample Location

The guidance provided in Subsection 3.3.5 should be followed when identifying sample locations in wetland sites. If sampling in a treed wetland (e.g., swamps and fens) flagging tape around existing trees or shrubs can be used to identify sampling sites; however, if sampling in an open area or an area that is underwater, stakes (wooden, metal, or PVC) are the most appropriate site markers. Stakes must be long enough that they can be driven into the substrate and still be higher than the high water level. Over time (with seasonal changes) stakes may be moved by water flow and/or ice movement. For this reason, hand-held GPS systems should be used in combination with site markers and a detailed site description (Subsection 3.3.8) to identify sample locations at a wetland site.

6.5.2 Sampling Wetland Soils

6.5.2.1 Access

There are a number of considerations to take into account when sampling wetland soils. The first and most important consideration is the access to the site and, more specifically, to the soil to be sampled. Access to the sampling site can be difficult, as wetlands are often surrounded by dense bush and/or water and access to samples on site can be challenging due to the variability of water and firm ground within wetlands. Wetland soils could be under metres of water or could even be floating on the water; either condition makes sampling difficult and potentially dangerous. Due to poor accessibility, inability to use heavy equipment, and/or the sensitivity of wetland ecosystems to disturbance, it is difficult, and often impossible, to collect large sample volumes of wetland soils. This can be a significant limitation in the amount of soil available for biological testing as wetland soils

are typically comprised primarily of water (e.g., up to 85%) resulting in significant loss of soil mass and volume upon drying. Therefore early consultation with the soil *ecotoxicology* laboratory is recommended in order to determine the number and volume of soil samples to collect to ensure sufficient soil for testing, and to adjust the experimental design of the site assessment if necessary.

6.5.2.2 Wetland Variability

Another important consideration when sampling wetland soils is wetland variability. Wetlands, and especially peatlands, exhibit appreciable micro-environmental variation. Perhaps the most important of these in bogs and fens is the micro-topographic variation created by vegetative growth. Various sphagnidae grow in hummocks, interspersed with small depressions and flooded pools. Within bogs and fens, different species of sphagnidae occupy the tops versus the bases of hummocks, and most plants show some degree of zonation relative to proximity to the seasonal or annual water table. Such hummocks also result in a distributional mosaic of various keystone vascular plants, from smaller woody shrubs such as dwarf birch or Labrador tea to black spruce seedlings. The strong micro-topographic relief is almost invariably accompanied by lateral variation in hydrological characteristics and degree of saturation of near-surface soils. It also has implications for smaller scale variation in contaminant distribution.

The high degree of spatial variability needs to be taken into consideration when sampling wetland soils. Typically, the size of sampling units is relatively small; for example, a 10- to 20-cm diameter peat corer sampled either singly, or as multiple core samples pooled into a composite sample from an area ≥ 1 m^2. The upper surface of the substrate to be collected should be arbitrarily established as the top of hummocks within a few metres of the area of interest.

Vertical changes in soil characteristics in peatlands tend to be gradational rather than highly discrete and profound. At the surface are actively growing bryophytes and vascular plants. Beneath this layer, it is often challenging to identify where active bryophyte growth ends and major accumulation of detrital organic matter commences. The depth of accumulated organics in peatlands can vary from a few tens of centimetres to ten metres or more. Changes in the degree of water saturation within the soil profile also tend to be highly gradational. The depth to water table can readily be identified by measuring from the surface of the peatland to the depth of standing water in a core hole or other excavation. Above this depth, however, peat soils also exhibit a high degree of saturation as they are very hydrophilic and have a high wicking potential. The transition from most saturated to least saturated soil can occur over small vertical distances (e.g., a few cm) or very large vertical distances (e.g., > 0.5 m).

Very little is known about the characteristics, diversity, or productivity of microorganisms or mesofauna with depth from surface in Canadian wetland systems. Preliminary results for boreal peatland systems; however, suggest that mesofaunal abundance (dominated by soil mites and springtails) is greatest within the fibric zone immediately adjacent to actively growing moss, and is very limited at soil depths below the water table (D. Bright, UMA Engineering, *pers. com.*, 2010). Wetland soils are likely to exhibit a strong vertical gradient in redox potential, with the oxidized zone occupying the upper few centimetres or more depending on local conditions. Contaminants may exhibit higher concentrations within the fully saturated zone of wetlands relative to the oxidized (unsaturated) zone depending on their physicochemical and fate characteristics.

6.5.2.3 Sampling for Toxicity Testing

Soil sampling in peaty wetlands can be done by coring or by cutting sods of peat using a serrated knife. Some specialized samplers for wetlands (particularly peat) have been developed by researchers (Zoltai, 1978; Cahoon et al., 1996; Buttler et al., 1998; LaForce et al., 2000; Reinhardt et al., 2000) (Table G.1; Appendix G). Sheppard et al. (1993) describes several of these tools and

techniques, many of which use corers that have been designed with very sharp and/or serrated cutting tips to enable coring through peat. Sampling peaty wetlands using peat harvesting techniques is not recommended as it is destructive to wetland habitat (Subsection 6.4).

Shallow open water, marsh, and swamp wetland soils do not have the organic/peaty accumulation typical of soils from bogs and fens; instead these types of wetlands typically have nutrient-rich mineral soils. Since these wetlands are subject to large seasonal water level fluctuations, it is recommended to plan sampling events during the seasonal low water level periods whenever possible. If sampling occurs during a low water level period, these soils can be sampled using the tools and techniques described earlier (Subsection 3.6.5). If sampling a submerged wetland site is necessary or unavoidable, it is recommended that the tools and techniques designed for sediment sampling are used; the reader is directed to the guidance on sampling sediments for biological testing that can be found in Environment Canada EPS 1/RM/29 (EC, 1994).

6.5.2.4 Sampling for Microbial Testing

Microbial communities can be the most dominant community in some wetlands; therefore, microbial testing of wetland soils might be an important component of a contaminated wetland site assessment (Mueller *et al.*, 2003). The following should be considered when sampling wetland soils for microbial testing: anaerobic sub-surface soil conditions, microbial-vegetative community interactions, seasonal water level changes, and, soil and pore-water chemistry.

Wetland soils can be aerobic or anaerobic. Anaerobic conditions can form beneath the soil surface and often the community composition and health of microorganisms that thrive in this environment might be of interest in a contaminated site assessment (Prietzel *et al.*, 2009). When testing anaerobic soil samples care must be taken to ensure anaerobic conditions are maintained throughout sample collection, transport, preparation, and testing. There are some tools and techniques designed to sample cores anaerobically. For example, samples can be collected using a steel auger, quickly loaded into glass jars, and then immediately sealed and purged with pure argon through a hole in the lid (the hole is stopped once the purge has been completed) (Prietzel *et al.*, 2009). Piston coring devices and vacuum-chamber transporters to sample 20-cm intact cores anaerobically and maintain their anaerobic state during transportation have also been used (LaForce *et al.*, 2000).

Wetlands are very sensitive to disturbance and the chemistry of soil samples and soil pore-water can change during transportation, storage, and preparation (e.g., drying). Changes to soil pH, metal speciation, and bioavailability, other contaminant bioavailability, nutrient bioavailability, and oxidation-reduction potential can result in changes to microbial communities in soil samples (Weider and Lang, 1986; Zoltai, 1995; LaForce *et al.*, 2000; OldeVenterink *et al.*, 2002). Care should be taken when sampling, transporting, storing, and preparing wetland soil samples to minimize changes to the soil from normal field conditions as much as possible.

Soil pore-water strongly influences microbial communities that live in wetland environments. As a result, both soil and soil pore-water should be chemically characterized when wetland soils are sampled for microbial testing.

6.5.3 Transport of Wetland Soils

In general, the guidance provided in Subsection 3.8 is applicable for wetland sample transport. Samples may be wet and therefore can be very heavy. The amount of sample per sample container should be distributed so that containers are not too heavy to lift. If heavy samples are being shipped they should be properly labelled to indicate their weight. Sample containers must be waterproof. All wetland samples should be kept on ice or in cool conditions (e.g., $4 \pm 2^{\circ}C$) during

transport. When shipping anaerobic samples special consideration should be given to ensure that the anaerobic conditions are maintained during transportation. The duration of transportation should be expedited whenever possible.

6.5.4 Storage of Wetland Soils

The storage guidance provided in Subsection 3.9.2 applies to wetland soil samples. In particular, wetland soils should be stored in a cool (e.g., 4 ± 2°C), dark location and should not be frozen. The duration of storage will depend on the wetland soil type, type of contaminant, and the testing to be conducted. Bogs and poor fens (peat-forming wetlands) have slow rates of decomposition due to low pH and the fact that nutrients are bound in the peat; therefore, soils from these wetlands will remain unchanged from field conditions for longer storage periods (preferably < 2 weeks, but up to 6 weeks) than soils from marshes, swamps, and rich fens. Nutrient availability in marshes, swamps, and rich fens soils is greater, and therefore these soils have a higher rate of production and decomposition. Soil characteristics such as pH, microbial communities, metal speciation and bioavailability and, depending on the contaminant, the bioavailability of other contaminants can change rapidly in these types of wetland soils (Zoltai and Vitt, 1995). Since these changes can occur rapidly it is recommended that these soils be tested immediately upon arrival at the ecotoxicology laboratory or to minimize the storage time as much as possible (preferably < 72 hours, but up to 2 weeks). Communication between sampling and laboratory testing personnel should be done at the early stages of the project so that sampling and testing activities are coordinated to eliminate or reduce the storage time of wetland soils.

Anaerobic samples should remain undisturbed (samples should remain unopened) until testing, and aerobic samples should be aerated if necessary to preserve the aerobic state until testing. To maintain aerobic conditions for soil samples, unseal the sample container and allow the soil to passively dry to a workable moisture content (Subsection 3.10.3.1) but not completely dry. If the soil is completely saturated but not covered with water then the soil should be air-dried as soon as possible upon arrival, otherwise anaerobic conditions will quickly develop. If the soil sample arrives submerged but was not collected submerged (e.g., if it was collected frozen, or if the saturated soil lost its structure and the soil moisture collected on the surface) then the water on the soil surface soil should be decanted and the soil air-dried as soon as possible to avoid the development of anaerobic conditions (Subsection 3.10.3.1). If the soil sample arrives submerged because it was collected submerged then it should not be disturbed prior to testing.

6.5.5 Preparation of Wetland Soils

The handling and preparation of wetland soils should be kept to a minimum and soil manipulation is not recommended. If any preparation is required in order to test wetland soil samples, the soil will no longer reflect field conditions and the testing results might not meet study DQOs. This is true regardless of wetland type. It is strongly recommended that wetland soil samples be fully chemically characterized before and after any soil preparation in order to quantify any changes due to preparation (Subsection 6.5.6).

Drying and/or drying and re-wetting wetland soils will alter the soil chemistry; therefore, careful consideration of the study objectives and/or DQOs should be given to determine the necessity of drying wetland soils before testing (Das and Maiti, 2008). The rapid rate of microbial production and decomposition in marsh and swamp soils make them very sensitive to changes in soil moisture content, which can result in shifts in nutrient bioavailability and other chemical and biological characteristics. However, depending on the study objectives, drying and re-wetting marsh and swamp soils to mimic natural fluctuations in moisture content (seasonal water variation) might be appropriate.

When preparing wetland soils for testing, careful attention should be paid to potential

hysteresis effects, particularly for peaty soils. Hysteresis is the change in soil properties such as gas diffusivity, water retention, air content and conductivity observed in soils when re-wetted after being dried (Caron *et al.* 2008). For example, peaty wetland soils can become hydrophobic when dried. Hysteresis, when it occurs, makes re-wetting soil samples very difficult and will cause changes to the physical properties of the soil (Caron *et al.* 2008). Water holding capacity cannot be measured using the dry weight method (EC 2004a, 2005a, 2007a) with these types of soils. An appropriate method for determining the water holding capacity of peaty soils is described in Parent and Caron (1993).

6.5.6 Physical and Chemical Characterization

Recommendations provided in Subsection 3.11 should be followed for the physical and chemical characterization of wetland soils along with the following additional guidance.

Concentrations of both ammonia and sulphide can reach sufficiently high levels under some biological test conditions to influence test results and levels of both substances should be measured prior to testing. Changes in the reducing conditions of wetland soils can also result in changes to the speciation or bioavailability of contaminants (Pascoe *et al.*, 1993; Das and Maiti, 2008; Prietzel *et al.*, 2009). In some wetlands, much of the plant life is dependent on microbial activity (e.g., denitrification). Changes in wetland soil chemistry (e.g., pH, redox potential, nutrient levels) and microbial community might bias the results of biological tests such that the tests no longer reflect field conditions as closely as they could. As with terrestrial soils, it is important to know how a contaminant(s) behaves under both field and laboratory conditions.

6.5.7 Test Species Selection

Due to the semi-aquatic nature of wetlands, current standardized test species would not generally be appropriate for biological testing of wetland soils. Adjusting the soil to accommodate the needs of currently standardized test species (e.g., pH adjustment, drying) is not appropriate and will introduce artefacts into the toxicity test data. Instead, test species should be selected that are naturally adapted to, or tolerant of, the conditions found in the type of wetland that is being assessed. Currently the most common type of wetland site soil assessment evaluates indigenous microbial communities (Steevens *et al.*, 1998; Mueller *et al.*, 2003), and less commonly, indigenous plant species (Carbonnell *et al.*, 1998). Soil invertebrates are not likely to be found in saturated wetland soils (e.g., those submerged below the water table). Some characterization of terrestrial peatland arthropods have been completed through ecological surveys (Finnamore and Marshall, 1994). Environment Canada is currently developing terrestrial toxicity test methods with test species adapted to boreal forest conditions including wetland plant and invertebrate species (EC, 2009).

References

AAFC (Agriculture and Agri-Food Canada). 1998. The Canadian system of soil classification. 3rd ed. Soil Classification Working Group, Agriculture and Agri-Food Canada Publication 1646, National Research Council (NRC) Research Press, Ottawa, pp. 187.

AAFC (Agriculture and Agri-Food Canada). 2007. Soil Landscapes of Canada Version 3.1.1. August 2009, http://sis.agr.gc.ca/cansis/nsdb/slc/v3.1.1/intro.html

AC (Agriculture Canada). 1977. Soils of Canada. Volume 1 Soil Report. The Canada Soil Survey Committee and The Soil Research Institute. Clayton, J.S., Ehrlich, W.A, Cann, D.B., Day, J.H., and Marshall, I.B. Printing and Publishing Supply and Services Canada. ISBN 0-660-00502-6. pp. 343.

AC (Agriculture Canada). Research Branch. 1981. A soil mapping system for Canada: Revised. Mapping System Working Group. Land Resource Institute. Contribution Number 142. pp. 94.

AE (Alberta Environment). 2001. Salt contamination assessment and remediation guidelines. Environmental Science Division, Environmental Service, Pub. No. T/606. May 2001.

AE (Alberta Environment). 2007a. Alberta Tier 2 Soil and Groundwater Remediation Guidelines. June 2007 ISBN: 978-0-7785-6308-2 (On-line Version).

AE (Alberta Environment). 2007b. Tier 2 Eco-contact Guideline Derivation Protocol. July 13 2007 Draft. ISBN: 978-0-7785-6753-0 (On-line Version).

AE (Alberta Environment). 2007c. Alberta Tier 1 Soil and Groundwater Remediation Guidelines. June 2007 ISBN: 978-0-7785-6306-8 (On-line Edition).

AE (Alberta Environment). 2008. Alberta Environmental Site Assessment Guidelines, Draft June 2008. ISBN: 978-0-7785-7371-5 (On-line Edition).

AESA (Alberta Environmentally Sustainable Agriculture Program). 2001. AESA soil quality benchmark study – soil organic matter. Factsheet FS2001-ISQ, 4 p., ASEA Soil Quality Program, Conservation and Development Branch, Alberta Agriculture, Food and Rural Development, Edmonton, Alberta.

Alberta Sustainable Resource Development. 2003. Ecological land survey site description manual. 2nd edition. Resource Data Branch. Strategic Corporate Services Division. Pub No: T/036. ISBN: 2819-7.

Anaka, A, Wickstrom, M and Siciliano, SD. 2008. Increased sensitivity and variability of phytotoxicity responses in arctic soils to a reference toxicant, boric acid. Environmental Toxicology and Chemistry, 27(3): 720-26.

Archibald, JH, and Beckingham, JD. 1996. Field guide to ecosites of northern Alberta. Special Report 5, Canadian Forest Service. Northwest Region, Northern Forestry Centre.

ASTM (American Society for Testing and Materials). 1993. Standard guide for conducting a terrestrial soil-core microcosm test. E 1197. American Society for Testing and Materials, West Conshohocken, PA.

ASTM (American Society for Testing and Materials). 2001. Standard guide for conducting laboratory soil toxicity test with the nematode *Caernorhabditis elegans*. E 2172-01. American Society for Testing and Materials, West Conshohocken, PA.

ASTM (American Society for Testing and Materials). 2004. Standard guide for conducting laboratory soil toxicity test or bioaccumulation tests with the lumbricid earthworm *Eisenia fetida* and the enchytraeid potworm *Enchytraeus albidus*. E1676-04. American Society for Testing and Materials, West Conshohocken, PA.

ASTM (American Society for Testing and Materials). 2005. Guide for selection of soil and rock sampling devices with drill rigs for environmental investigations. D6169. American Society for Testing and Materials, West Conshohocken, PA.

ASTM (American Society for Testing and Materials). 2006a. Standard classification of soil for engineering purposes. (Unified Soil Classification System). D2487. American Society for Testing and Materials, West Conshohocken, PA.

ASTM (American Society for Testing and Materials). 2006b. Standard guide for soil sampling from the vadose zone. D4700-91. American Society for Testing and Materials, West Conshohocken, PA.

ASTM (American Society for Testing and Materials). 2006c. Standard Guide for Sampling Waste Piles. D6009-96. American Society for Testing and Materials, West Conshohocken, PA.

ASTM (American Society for Testing and Materials). 2008a. Standard practice for decontamination of field equipment used at nonradioactive waste sites. D5088-90. American Society for Testing and Materials, West Conshohocken, PA.

ASTM (American Society for Testing and Materials). 2008b. Test method for penetration test (SPT) and split-barrel sampling of soils. D1586-08. American Society for Testing and Materials, West Conshohocken, PA.

ASTM (American Society for Testing and Materials). 2008c. Practice for thin-walled tube sampling of soils for geotechnical purposes. D1587. American Society for Testing and Materials, West Conshohocken, PA.

ASTM (American Society for Testing and Materials). 2008d. Standard practice for using hollow-stem augers for geotechnical exploration and soil sampling. D6151. American Society for Testing and Materials, West Conshohocken, PA.

ASTM (American Society for Testing and Materials). 2009a. Practice for soil investigation and sampling by auger borings. D1452. American Society for Testing and Materials, West Conshohocken, PA.

ASTM (American Society for Testing and Materials). 2009b. Practice for thick wall, ring-lined, split barrel, drive sampling of soils. D3550. American Society for Testing and Materials, West Conshohocken, PA.

Athey, LA, Thomas, JM and Skalski, JR. 1987. Role of acute toxicity bioassays in the remedial action process at hazardous waste sites. United States Environmental Protection Agency, Environmental Research Laboratory, Corvallis, OR 97333, EPA/600/8-87/044.

Bailey, VL, Bolton Jr H and Smith, JL. 2008. Substrate-inducted respiration and selective inhibition as measures of microbial biomass in soils. *In*: Carter, MR and Gregorich, EG (Eds). 2008. Soil sampling and methods of analysis, 2nd Edition. Canadian Society of Soil Science, CRC Press Taylor and Francis Group, Boca Raton, FL, pp. 515-526.

BC MELP (B.C. Ministry of Environment, Lands and Parks). 1995. Technical guidance on Contaminated Sites: Statistical Criteria for Characterizing a Volume of Contaminated Material.

BC Ministry of Forests and BC Ministry of Environment, Lands and Parks. 1998. Field manual for describing terrestrial ecosystems. Land Management Handbook Number 25. ISSN 0229-1622.

Bélanger, N and Van Rees, KJC. 2008. Sampling forest soils. *In*: Carter, MR and Gregorich, EG (Eds). 2008. Soil sampling and methods of analysis, 2nd Edition. Canadian Society of Soil Science, CRC Press Taylor and Francis Group, Boca Raton, FL.

Bélanger, N, Paré, D and Hendershot, WH. 2008. Determining nutrient availability in forest soils. *In*: Carter, MR and Gregorich, EG (Eds). 2008. Soil sampling and methods of analysis, 2nd Edition. Canadian Society of Soil Science, CRC Press Taylor and Francis Group, Boca Raton, FL.

Bellhouse, DR and Sutradhar, BC. 1988. Variance estimation for systematic sampling when autocorrelation is present. The Statistician. 37:327-332.

Beyaert, RP and Fox, CA. 2008. Assessment of soil biological activity. *In*: Carter, MR and Gregorich, EG (Eds). 2008. Soil sampling and methods of analysis, 2nd Edition. Canadian Society of Soil Science, CRC Press Taylor and Francis Group, Boca Raton, FL, pp. 527-545.

Blaylock, AD. 1994. Soil salinity, salt tolerance and growth potential of horticultural and landscape plants. University of Wyoming Cooperative Extension Service, Department of Plant, Soil, and Insect Sciences, College of Agriculture, B-988, February 1994.

Bombardier, M. 2004. Soil microcosm test systems for assessing the potential environmental effects of chemical or microbial substances: A review. Method Development and Applications Section, Environmental Technology Centre, Environment Canada, Ottawa, ON, pp. 109, Appendices.

Breckenridge, RP and Crockett, AB. 1995. Determination of background concentrations of inorganics in soils and sediments at hazardous waste sites. EPA/540/S-96/500.

Buttler, A, Grosvernier, P and Matthey, Y. 1998. A new sampler for extracting undisturbed surface peat cores for growth pot experiments. New Phytologist, 140(2):355-360.

Cahoon, DR, Lynch, JC and Knaus, RM. 1996. Improved cryogenic coring device for sampling wetland soils. Journal of Sedimentary Research. 66:1025-1027.

Campbell, DA, Cole, CA and Brooks, RP. 2002. A comparison of created and natural wetlands in Pennsylvania, USA.

Carbonell, AA, Aarabi, MA, DeLaune, RD, Gambrell, RP, Patrick Jr WH. 1998. Arsenic in wetland vegetation: Availability, phytotoxicity, uptake and effects on plant growth and nutrition.

Caron, J, Elrick, DE, Michel, JC, and Naasz, R. 2008. Physical properties of organic soils and growing media: Water and air storage and flow dynamics. *In*: Carter, MR and Gregorich, EG (Eds). 2008. Soil sampling and methods of analysis, 2nd Edition. Canadian Society of Soil Science, CRC Press Taylor and Francis Group, Boca Raton, FL, pp. 885-912.

Carter, MR (Ed.). 1993. Soil Sampling and Methods of Analysis. Canadian Society of Soil Science, CRC Press Taylor and Francis Group, Boca Raton, FL.

Carter, MR, and Gregorich, EG (Eds.). 2008. Soil Sampling and Methods of Analysis, 2nd Edition. Canadian Society of Soil Science, CRC Press Taylor and Francis Group, Boca Raton, FL.

CCME (Canadian Council of Ministers of the Environment). 1993a. Guidance manual on sampling, analysis and data management for contaminated sites – Volume I. CCME Subcommittee on Environmental Quality Criteria for Contaminated Sites, the National Contaminated Sites Remediation Program, March 1993, Winnipeg Manitoba.

CCME (Canadian Council of Ministers of the Environment). 1993b. Guidance manual on sampling, analysis and data management for contaminated sites – Volume II. CCME Subcommittee on Environmental Quality Criteria for Contaminated Sites, the National Contaminated Sites Remediation Program, March 1993, Winnipeg Manitoba.

CCME (Canadian Council of Ministers of the Environment). 1996a. A framework for ecological risk assessment: General guidance. PN 1195 CCME Subcommittee on Environmental Quality Criteria for Contaminated Sites, the National Contaminated Sites Remediation Program, March 1996, Winnipeg Manitoba.

CCME (Canadian Council of Ministers of the Environment). 1996b. Guidance manual on developing site-specific soil quality remediation objectives for contaminated sites in Canada. PN 1197 CCME Subcommittee on Environmental Quality Criteria for Contaminated Sites, the National Contaminated Sites Remediation Program. March 1996, Winnipeg Manitoba.

CCME (Canadian Council of Ministers of the Environment). 1997. A framework for ecological risk assessment: Technical appendices. PN 1195 CCME Subcommittee on Environmental Quality Criteria for Contaminated Sites, the National Contaminated Sites Remediation Program, March 1997, Winnipeg Manitoba.

CCME (Canadian Council of Ministers of the Environment). 1999. Protocol for the derivation of Canadian sediment quality guidelines for the protection of aquatic life. CCME EPC-98E http://ceqg-rcqe.ccme.ca/download/en/226/

CCME (Canadian Council of Ministers of the Environment). 2006. A protocol for the derivation of environmental and human health soil quality guidelines. PN 1332 ISBN-10 1-896997-45-7 PDF

Certini, G, Campbell, CD, and Edwards, AC. 2004. Rock fragments in soil support a different microbial community from the fine earth. Soil Biology and Biochemistry, 36(2004):1119-1128.

CFEM (Canadian Foundation Engineering Manual). 2007. 4th Edition, Bitech Publishers Ltd.

CSA (Canadian Standards Association). 2001. Phase I environmental site assessment. Z768-01 (R2006).

Cochran, WG. 1977. Sampling Techniques. 3rd Ed. John Wiley & Sons, Toronto.

Compton, R. 1985. Geology in the field. John Wiley and Sons Inc. New York, New York. ISBN: 0-471-82902-1.

Cordy, C. 1993. An extension of the Horvitz-Thompson theorum to print sampling from a continuous universe. Statistics and Probability Letters. 18(5): 353-362.

Crepin, J and Johnson, RL. 1993. Soil sampling for environmental assessment. *In*: Carter, M.R. (Ed.), 1993. Soil Sampling and Methods of Analysis. Canadian Society of Soil Science, CRC Press Taylor and Francis Group, Boca Raton, FL, pp. 5-18.

Cressie, NC. 1993. Statistics for spatial data. John Wiley & Sons, Toronto.

Culley, JLB. 1993. Density and compressibility. *In*: Carter, MR (ed), 1993. Soil Sampling and Methods of Analysis. Canadian Society of Soil Science, CRC Press Taylor and Francis Group, Boca Raton, FL, pp. 529-539.

Dalpé, Y and Hamel, C. 2008. Arbuscular mycorrhizae. In: Carter, MR and Gregorich, EG (Eds). 2008. Soil sampling and methods of analysis, 2nd Edition. Canadian Society of Soil Science, CRC Press Taylor and Francis Group, Boca Raton, FL, pp. 355-377.

Danke, HV. 1981. Arctic Arthropods – A review of systematic and ecology with particular reference to the North American fauna. Entomological Society of Canada, Ottawa.

Das, M and Maiti, SK. 2008. Comparison between availability of heavy metals in dry and wetland tailing of an abandoned copper tailing pond. Environmental Monitoring Assessment, 137:343-350.

Dayton, EA, Basta, NT, Payton, ME, Bradham, KD, Schroder, JL and Lanno, RP. 2006. Evaluating the contribution of soil properties to modifying lead phytoavailability and phytotoxicity. Environmental Toxicology and Chemistry, 25(3): 719-725.

Department of Environment (UK). 1994. Guidance on Preliminary Site Inspection of Contaminated Land. CLR No.2 Volume 2 (of 2). Prepared by Environmental Research Centre Ltd. http://www.eugris.info/envdocs/CLR0202.pdf

Dirven-van Breemen, EM, Mesman, M, Rutgers, M, Jensen, J, Sorokin, N, Bierkens, J, Ter Laak, T, Erlacher, L and Bogolte, T. 2006. The case study Skagen, Denmark. *In*: Jensen, J and Mesman, M (Eds). Ecological risk assessment of contaminated land: Decision support for site specific investigations. RIVM report number 711701047, ISBN 978-90-6960-138-0, pp. 117-126.

Drury, CF, Hart, SC and Yan, XM. 2008a. Nitrification techniques for soils. *In*: Carter, MR and Gregorich, EG (Eds). 2008. Soil sampling and methods of analysis, 2nd Edition. Canadian Society of Soil Science, CRC Press Taylor and Francis Group, Boca Raton, FL, pp. 495-513.

Drury, CF, Myrold, DD, Beauchamp, EG and Reynolds, WD. 2008b. Denitrification techniques for soils. *In*: Carter, MR and Gregorich, EG (Eds). 2008. Soil sampling and methods of analysis, 2nd Edition. Canadian Society of Soil Science, CRC Press Taylor and Francis Group, Boca Raton, FL, pp. 471-493.

Eberhardt LL, and Tomas, JM. 1991. Designing Environmental Field Studies. Ecol. Mon. 61(1):53-73.

Echevarria, G, Massoura, ST, Sterckman, T, Becquer, T, Schwartz and Morel, JL. 2006. Assessment and control of the bioavailability of nickel in soils. Environmental Toxicology and Chemistry, 25(3):643-651.

EC (Environment Canada). 1994. Guidance document on collection and preparation of sediments for physicochemical characterization and biological testing. Technology Development Directorate, Environment Canada, Ottawa, ON, EPS 1/RM/29, December 1994.

EC (Environment Canada). 1999. Guidance document on application and interpretation of single-species tests in environmental toxicology. Report EPS 1/RM/34, December 1999. Method Development and Applications Section, Environmental Technology Centre, Environment Canada, Ottawa, Ontario.

EC (Environment Canada). 2002a. Technical assistance bulletin #1. Jar Headspace Analytical Screening Procedures. Environment Canada Ontario Region, Environmental Protection Branch, Environmental Contaminants and Nuclear Programs Division. www.on.ec.gc.ca/pollution/ccnpd/tabs/tab01-e.html

EC (Environment Canada). 2002b. Technical assistance bulletin #6. Alternative Field Screening Methods. Environment Canada Ontario Region, Environmental Protection Branch, Environmental Contaminants and Nuclear Programs Division. www.on.ec.gc.ca/pollution/ccnpd/tabs/tab06-e.html

EC (Environment Canada). 2002c. Metal Mining Guidance Document for Aquatic Environmental Effects Monitoring, June 2002.

EC (Environment Canada). 2003a. Toxicity test methodologies for assessing the impacts of contaminant mixtures in soil using terrestrial species of ecological relevance to Canadian soil systems. Workshop Proceedings, February 19, 20 and 21, 2003. Environment Canada's Pacific Environmental Sciences Centre, Vancouver, British Columbia.

EC (Environment Canada). 2003b. Toxicity test methodologies for assessing the impacts of contaminant mixtures in soil using terrestrial species of ecological relevance to Canadian soil systems: Workshop Appendices, February 19, 20 and 21, 2003. Environment Canada's Pacific Environmental Sciences Centre, Vancouver, British Columbia.

EC (Environment Canada). 2004a. Biological Test Method: Tests for Toxicity of Contaminated Soil to Earthworms (*Eisenia andrei*, *Eisenia fetida*, or *Lumbricus terrestris*). Report EPS 1/RM/43, June 2004. Method Development and Applications Section, Environmental Technology Centre, Environment Canada, Ottawa, Ontario.

EC (Environment Canada). 2004b. Pulp and Paper EEM Guidance Document. May-04.

EC (Environment Canada). 2005a. Biological Test Method: Test for Measuring Emergence and Growth of Terrestrial Plants Exposed to Contaminants in Soil. Report EPS 1/RM/45, February 2005. Method Development and Applications Section, Environmental Technology Centre, Environment Canada, Ottawa, Ontario.

EC (Environment Canada). 2005b. Terrestrial ecozones of Canada. State of the Environment Infobase. http://www.ec.gc.ca/soer-ree/English/Vignettes/Terrestrial/terr.cfm

EC (Environment Canada). 2007a. Biological Test Method: Test for Measuring Survival and Reproduction of Springtails Exposed to Contaminants in Soil. Report EPS 1/RM/47 September 2007. Method Development and Applications Section, Environmental Technology Centre, Environment Canada, Ottawa, Ontario.

EC (Environment Canada). 2007b. Sampling of Contaminated Soils in the Swan Hills and Peace River Areas of Alberta. Prepared for: Method Development and Applications Section, Environmental Technology Centre, Environment Canada, Ottawa, Ontario. Prepared by: EcoDynamics Consulting Inc., Prince Albert Saskatchewan, February 2 2007, pp. 10.

EC (Environment Canada). 2008. Canada-Ontario decision-making framework for assessment of Great Lakes contaminated sediment. Peter Chapman (Golder Associates Ltd.) and COA (Canada Ontario Agreement) Sediment Task Group. ISBN 978-0-662-46147-0 http://publications.gc.ca/collections/collection_2010/ec/en164-14-2007-eng.pdf

EC (Environment Canada). 2009. PERD Status Report 2009. Prepared by: Biological Assessment and Standardization Section, Environment Canada, December 2009.

EC and SRC (Environment Canada and Saskatchewan Research Council). 2007. Validation of toxicology test methods for assessing petroleum hydrocarbon and brine spills in boreal forest soils. Prepared for: Environmental Research Advisory Council, Canadian Association of Petroleum Producers. Prepared by: Biological Methods Division Environment Canada and Environment and Forestry Division Saskatchewan Research Council, October 2007.

Ehrlichmann, H, Eisenträger, A, Möller, M and Dott, W. 1997. Effect of storage conditions of soil on ecotoxicological assessment. International Biodeterioration and Biodegradation, 39(1):55-59.

ESG (ESG International Inc.). 2000. Assessment of the biological test methods for terrestrial plants and soil invertebrates: Metals. Prepared for Method Development and Applications Section, Environmental Technology Centre, Environment Canada, Ottawa, ON. pp. 40 (Appendices).

Fierer, N and Schimel, JP. 2002. Effects of drying-rewetting frequency on soil carbon and nitrogen transformations. Soil Biology and Biochemistry, 34:777-787.

Finnamore, AT, and Marshall, SA. (Eds.) 1994. Terrestrial arthropods of peatlands, with particular reference to Canada. Memoirs of the Entomological Society of Canada, 169:3-289.

Flores, AL, Martínez, LI and Ferrer, CM. 2003. Systematic sample design for the estimation of spatial means. Environmetrics, 14:45–61.

Forbes, VE and Forbes, TL. 1994. Ecotoxicology in theory and practice. Chapman & Hall Ecotoxicology Series 2: London.

Fränzle, O and Kuhnt, G. 1994. Fundamentals of representative soil sampling. In: Kuhnt, G., Muntau, H. Euro-Soils – Identification, collection, treatment, characterization. Environment Institute, European Commission, Joint Research Centre, Ispra, Italy. Special Publication No. 1.94.60. pp. 11-29.

Garcia, MV. 2004. Effects of pesticides on soil fauna: Development of ecotoxicological test methods for tropical regions. Ecology and Development Series No. 19. Zentrum fuer Entwicklungsforschung, University of Bonn, Germany.

Garcia, MV, Römbke, J, de Brito, MVT and Scheffczyk, A. 2008. Effects of three pesticides on the avoidance behaviour of earthworms in laboratory tests performed under tropical conditions. Environmental Pollution, 153:450-456.

Gaudet, C, Bright, D, Adare, K and Potter, K. 2001. A rank-based approach for deriving Canadian soil and sediment quality guidelines. *In*: The use of species sensitivity distributions (SSD) in ecotoxicology. Posthuma, L and Suter, GW (Eds.) CRC Press, Boca Raton, FL.

Germida, JJ. 1993. Cultural methods for soil microorganisms. *In*: Carter, MR (Ed), 1993. Soil sampling and methods of analysis. Canadian Society of Soil Science, CRC Press Taylor and Francis Group, Boca Raton, FL, pp. 263-275.

Gilbert, RO. 1987. Statistical methods for environmental pollution monitoring. Van Nostrand Reinhold, New York.

Green, RN, Trowbridge, RL and Klinka, K. 1993. Towards a taxonomic classification of humus forms, Forest Science Monograph 29. Society of American Foresters, Bethesda, MD.

Greene, JC, Bartels, CL, Warren-Hicks, WJ, Parkhurst, BR, Linder, GL, Peterson, SA and Miller, WE. 1989. Protocols for short term toxicity screening of hazardous waste sites. Prepared by the US Environmental Protection Agency Corvallis Environmental Research Laboratory Hazardous Waste Assessment Team, Corvallis, OR, Northrop Services Incorporated, Corvallis, OR and Kilkelly Environmental Associates, Raleigh, NC. Prepared for the United States Environmental Protection Agency Corvallis Environmental Research Laboratory, Office of Research and Development, Corvallis, OR 97333, EPA/600/3-88/029.

Hatscan. 2006. Handi-Guide to Alberta's OH&S Act, Regulation and Code. 2nd Edition. Hatscan Part 32. Excavating and Tunnelling. ISBN 978-1-895056-40-4.

Hausenbuiller, RL. 1985. Soil science – Principles and practices, 3rd Edition. W.C. Brown Publisher, Dubuque, Iowa.

The Hazardous Waste Consultant. 1992. Soil sampling and analysis – Practices and pitfalls. Special Feature in *The Hazardous Waste Consultant*. McCoy and Associates Inc. November/ December 1992.

Hoagland, DR and Arnon, DI. 1950. The water culture method of growing plants without soil. Circular No. 347, California Agricultural Experimental Station, University of California, Berkeley, CA.

Horvitz, DG, and Thompson, DJ. 1952. A generalization of sampling without replacement from a finite-universe. Journal of the American Statistical Association. 47: 663-685.

Hund-Rinke, K and Wiechering, H. 2001. Earthworm avoidance test for soil assessments; an alternative for acute and reproduction tests. Journal of Soils and Sediments, 1(1):15-20.

Hund-Rinke, K, Achazi, R, Rombke, J and Warnecke, D. 2003. Avoidance test with *Eisenia fetida* as indicator for the habitat function of soils: results of a laboratory comparison test. Journal of Soils and Sediments, 3(1):7-12.

Iachan, R. 1982. Systematic sampling: A critical review. International Statistical Review, 50:293-303.

INAC (Indian and Northern Affairs Canada). 2009. Guidelines for designing and implementing aquatic effects monitoring programs for development projects in the Northwest Territories: Overview Report and Technical Guidance Documents Volumes 1 to 6. Prepared by MacDonald Environmental Sciences Ltd. and Zajdlik and Associates Inc., in association with the Water Resources Division. Yellowknife, NT.

Isaaks, EH and Srivastava, RM. 1989. An introduction to applied geostatistics. Oxford University Press.

ISO (International Organisation for Standards). 1993a. 11268-1. Effects of pollutants on earthworms (*Eisenia fetida*). Part 1: Determination of acute toxicity using *artificial soil* substrate. Geneva, Switzerland.

ISO (International Organisation for Standards). 1993b. 11269-1. Soil quality: Determination of the effects of pollutants on soil flora. Part 1: Method for the measurement of inhibition of root growth. Geneva, Switzerland.

ISO (International Organisation for Standards). 1995. 11269-2. Soil quality: Determination of the effects of pollutants on soil flora. Part 2: Effects of chemicals on the emergence and growth of higher plants. Geneva, Switzerland.

ISO (International Organisation for Standards). 1997a. 14240-2. Determination of soil microbial biomass. Part 2: Fumigation extraction method. Geneva, Switzerland.

ISO (International Organisation for Standards). 1997b. 14240-1. Determination of soil microbial biomass. Part 1: Substrate induced respiration. Geneva, Switzerland.

ISO (International Organisation for Standards). 1997c. 14238. Biological methods. Determination of nitrogen mineralization and nitrification in soils and the influence of chemicals on these processes. Geneva, Switzerland.

ISO (International Organisation for Standards). 1998. 11268-2. Effects of pollutants on earthworms (*Eisenia fetida*). Part 2: Determination of effects on reproduction. Geneva, Switzerland.

ISO (International Organisation for Standards). 1999. 11267-2. Inhibition of reproduction of collembola (*Folsomia candida*) by soil pollutants. Geneva, Switzerland.

ISO (International Organisation for Standards). 2000. 13829. Water quality – Determination of the genotoxicity of water and waste water using the umu test. Geneva, Switzerland.

ISO (International Organisation for Standards). 2001. 10381-3. Soil quality -- Sampling -- Part 3: Guidance on safety. Geneva, Switzerland.

ISO (International Organisation for Standards). 2002a. 10381-1. Soil quality -- Sampling -- Part 1: Guidance on the design of sampling programmes. Geneva, Switzerland.

ISO (International Organisation for Standards). 2002b. 10381-2. Soil quality -- Sampling -- Part 2: Guidance on sampling techniques. Geneva, Switzerland.

ISO (International Organisation for Standards). 2002c. 10381-8. Soil quality -- Sampling -- Part 8: Guidance on sampling of stockpiles. Geneva, Switzerland.

ISO (International Organisation for Standards). 2002d. 16072. Laboratory methods for determination of microbial soil respiration. Geneva, Switzerland.

ISO (International Organisation for Standards). 2002e. 17155. Determination of abundance and activity of soil microflora using respiration curves. Geneva, Switzerland.

ISO (International Organisation for Standards). 2003a. 10381-4. Soil quality -- Sampling -- Part 4: Guidance on the procedure for investigation of natural, near-natural and cultivated sites. Geneva, Switzerland.

ISO (International Organisation for Standards). 2003b. 15799. Soil quality – Guidance on the ecotoxicological characterization of soils and soil materials. Geneva, Switzerland.

ISO (International Organisation for Standards). 2004a. 16387. Effects of pollutants on Enchytraeidae (*Enchytraeus* sp.). Determination of effects on reproduction and survival. Geneva, Switzerland.

ISO (International Organisation for Standards). 2004b. 15685. Determination of potential nitrification. Rapid test by ammonium oxidation. Geneva, Switzerland.

ISO (International Organisation for Standards). 2004c. 8692. Water Quality – Freshwater algal growth inhibition test with unicellular green algae. Geneva, Switzerland.

ISO (International Organisation for Standards). 2005a. 11074. Soil quality - Vocabulary. Geneva, Switzerland.

ISO (International Organisation for Standards). 2005b. 10381-5. Soil quality -- Sampling -- Part 5: Guidance on the procedure for the investigation of

urban and industrial sites with regard to soil contamination. Geneva, Switzerland

ISO (International Organisation for Standards). 2005c. 10381-7. Soil quality -- Sampling -- Part 7: Guidance on sampling of soil gas. Geneva, Switzerland.

ISO (International Organisation for Standards). 2005d. 17126. Determination of the effects of pollutants on soil flora. Screening test for emergence of lettuce seedlings (*Lactuca sativa* L.) Geneva, Switzerland.

ISO (International Organisation for Standards). 2005e. 20963. Effects of pollutants on insect larvae (*Oxythyrea funesta*). Determination of acute toxicity. Geneva, Switzerland.

ISO (International Organisation for Standards). 2005f. 22030. Biological methods. Chronic toxicity in higher plants. Geneva, Switzerland.

ISO (International Organisation for Standards). 2005g. 23753-1. Determination of dehydrogenase activity in soils. Part 1: Method using triphenyltetrazolium chloride (TTC). Geneva, Switzerland.

ISO (International Organisation for Standards). 2005h. 23753-2. Determination of dehydrogenase activity in soils. Part 2: Method using iodotetrazolium chloride. Geneva, Switzerland.

ISO (International Organisation for Standards) Draft. 2006a. 10381-6. Soil quality -- Sampling -- Part 6: Guidance on the collection, handling and storage of soil under aerobic conditions for the assessment of microbial processes, biomass and diversity in the laboratory. Geneva, Switzerland.

ISO (International Organisation for Standards). 2006b. 15952. Effects of pollutants on juvenile land snails (Helicidae). Determination of the effects on growth by soil contamination. Geneva, Switzerland.

ISO (International Organisation for Standards). 2006c. 23611-6. Soil quality. Sampling of soil invertebrates. Part 6: Guidance for the design of sampling programmes with soil invertebrates. Geneva, Switzerland.

ISO (International Organisation for Standards). 2006d. 10381-8. Soil quality. Sampling. Part 8: Guidance on the sampling of stockpiles. Geneva, Switzerland.

ISO (International Organisation for Standards). 2007a. 18512. Soil quality. Guidance on long and short term storage of soil samples. Geneva, Switzerland.

ISO (International Organisation for Standards). 2007b. 11348-1. Water quality – Determination of the inhibitory effect of water samples on the light emission of *Vibrio fischeri* (Luminescent bacteria test) – Part 1: Method using freshly prepared bacteria. Geneva, Switzerland.

ISO (International Organisation for Standards). 2007c. 10381-8. Soil quality. Sampling Part 8: Guidance on sampling of stockpiles. Geneva, Switzerland.

ISO (International Organisation for Standards). 2008a. 17512-1. Soil quality: Avoidance test for testing the quality of soils and effects of chemicals on behaviour. Part 1: Test with earthworms (*Eisenia andrei* and *Eisenia fetida*). Geneva, Switzerland.

ISO (International Organisation for Standards. 2008b draft. Water quality – Determination of the inhibition of dehydrogenase activity of *Arthrobacter globiformis* – Solid contact test using the redox dye resazurine. ISO 10871. Draft. Geneva, Switzerland.

IUPAC (International Union of Pure and Applied Chemistry). 2005. Terminology in soil sampling (IUPAC Recommendations 2005). Pure and Applied Chemistry, 77(5): 827-841.

Jackson, NE, Corey, JC, Frederick, LR, and Pick, JC-Jr. 1967. γ-Irradiation and the microbial population of soils at two water contents. In (2nd ed.) Soil Science Society of America Proceedings 31(1967), pp. 491-494.

Jänsch, S, Amorim, MJ, and Römbke, J. 2005. Identification of the ecological requirements of important terrestrial ecotoxicological test species. Environmental Reviews, 13:51-83.

Janssen RPT, Posthuma L, Baerselman R, Hollander RAD, van Veen RPM, and Peijnenburg WJGM. 1997. Equilibrium partitioning of heavy metals in Dutch field soils. II. Prediction of metal accumulation in earthworms. Environmental Toxicology and Chemistry, 16:2479-2488.

Jensen, J and Mesman, M (Eds). 2006. Ecological risk assessment of contaminated land: Decision support for site specific investigations. RIVM report number 711701047, pp. 136. ISBN 978-90-6960-138-0. pp. 136.

Jensen, J, Sorokin, N, Dirven-van Breemen, EM, Bogolte, T, Erlacher, E, Ehlers, C, Ter Laak, T, Hartnik, T, Bierkens, J, Rutgers, M and Mesman, M. 2006a. A triad-based selection of tools for site-specific assessment of ecological risk. *In*: Jensen, J and Mesman, M (Eds). Ecological risk assessment of contaminated land: Decision support for site specific investigations. RIVM report number 711701047, ISBN 978-90-6960-138-0, pp. 65-116.

Jensen, J, Mesman, M, Bierkens, J and Rutgers, M. 2006b. Principles and concepts in ecological risk assessment. *In*: Jensen, J and Mesman, M (Eds). Ecological risk assessment of contaminated land: Decision support for site specific investigations. RIVM report number 711701047, ISBN 978-90-6960-138-0, pp. 11-18.

Journel AG and Huijbregts, CJ. 1978. Mining Geostatistics. Academic Press, London.

Kerr, M and Stewart, AJ. 2003. Tolerance test of *Eisenia fetida* for sodium chloride. U.S. Department of Energy Journal of Undergraduate Research, 3:21-25.

Klinka, K, Green, RN and Trowbridge, RL. 1981. Taxonomic classification of humus forms in ecosystems of British Columbia. First approximation. British Columbia Ministry of Forest.

Knacker, T., Van Gestel, CAM, Jones, SE, Soares AMVM, Schallnab, H-J, Forste, B and Edwards, CA. 2004. Ring-testing and field validation of a Terrestrial Model Ecosystem (TME)- An instrument for testing potentially harmful substances: Conceptual Approach and Study Design. Ecotoxicology, 13:9-27.

Kools, SAE, Boivin, M-EY, Van Der Wurff, AWG, Berg, MP, Van Gestel, CAM, and Van Straalen, NM. 2009. Assessment of structure and function in metal polluted grasslands using Terrestrial Model Ecosystems. Ecotoxicology and Environmental Safety, 72:51-59.

Kuperman, R, Williams, G and Parmelee, R. 1998. Spatial variability in the soil food webs in a contaminated grassland ecosystem. Applied Soil Ecology, 9:509-514.

Kuperman, R, Knacker, T, Checkai, R and Edwards, C. 2002. Multispecies and multiprocess assays to assess the effects of chemicals on contaminated sites. *In*: Environmental analysis of contaminated sites. Sunahara, G, Renoux, A, Thellen, C, Gaudet, C and Pilon, A, Eds. John Wiley and Sons, Ltd. pp. 45-60.

Kuperman, RG, Checkai, RT, Simini, M, Phillips, CT, Kolakowski, JE and Kurnas, CW. 2005. Weathering and aging of 2,4,6-trinitrotoluene in soils increases toxicity to potworm *Enchytraeus crypticus*. Environmental Toxicology and Chemistry, 24(10):2509-2518.

Kuperman, RG, Checkai, RT, Simini, M, Phillips, CT, Kolakowski, JE and Kurnas, CW. 2006. Toxicities of dinitrotoluenes and trinitrobenzene freshly amended or weathered and aged in a sandy loam soil to *Enchytraeus crypticus*. Environmental Toxicology and Chemistry, 25(5):1368-1375.

LaForce, MJ, Hansel, CM, and Fendorf, S. 2000. Constructing simple wetland sampling devices. Soil Science Society of America Journal, 64:809-811.

Langdon CJ, Pierce TG Meharg AA, Semple KT. 2003. Interactions between earthworms and arsenic in soil: a review. Environmental Pollution, 124:361-373.

Lanno, R (Ed). 2003. Contaminated soils: From soil-chemical interactions to ecosystem management. Society of Toxicology and Chemistry, Pensacola, FL.

Linder, G, Ingham, E, Brandt, CJ and Henderson, G. 1992. Evaluation of terrestrial indicators for use in ecological assessments at hazardous waste sites. United States Environmental Protection Agency, Office of Research and Development, Washington, D. 20460. EPA/600/R-92/183.

Loibner, A, Jensen, J, Ter Laak, T, Celis, R and Hartnik, T. 2006. Sorption and ageing of soil contamination. *In*: Jensen, J and Mesman, M (Eds). Ecological risk assessment of contaminated land: Decision support for site specific investigations. RIVM report number 711701047, ISBN 978-90-6960-138-0, pp. 19-29.

Lotrario, JB, Stuart, BJ, Lam, T, Arands, RR, O'Connor, OA and Kosson, DS. 1995. Effects of sterilization methods of physical characteristics of soil: Implications for sorption isotherm analyses. Bulletin of Environmental Contamination and Toxicology, 54: 668-675.

Luttmerding, HA, Demarchi, DA, Lea, EC, Meidinger, DV, and Vold, T (Eds). 1990. Describing Ecosystems in the Field. Second edition. MOE

Manual 11. Province of British Columbia. Ministry of Environment, Lands and Parks, Ministry of Forests. ISSN 0821-0640. http://www.for.gov.bc.ca/hfd/pubs/docs/mr/mr074.pdf

Ma, Y, Lombi, E, Nolan, AL, and McLaughlin, MJ. 2006. Short-term natural attenuation of copper in soils: Effects of time, temperature and soil characteristics. Environmental Toxicology and Chemistry, 25(3): 652-658.

Mackenzie, WH and Moran, JR. 2004. Wetlands of British Columbia: A guide to identification. Research Branch, British Columbia Ministry of Forests, Victoria, BC. Land Management Handbook No. 52. *As cited in*: Risk based salinity guidelines for peatlands Phase 1: Literature review and gaps analysis. Canadian Association of Petroleum Producers, Peatland Salinity Working Group. Prepared by: UMA Engineering Ltd., 200-415 Gorge Road East, Victoria BC, V8T 2W1. Prepared for: the Canadian Association of Petroleum Producers, Peatland Salinity Working Group. Final Draft, May 2008.

Marshall, IB and Schut, PH 1999. A national ecological framework for Canada. Overview. Ecosystems Science Directorate, Environment Canada and Research Branch, Agriculture and Agri-Food Canada. http://sis.agr.gc.ca/cansis/nsdb/ecostrat/intro.html

Mason, BJ. 1992. Preparation of soil sampling protocols: Sampling techniques and strategies. United States Environmental Protection Agency, Office of Research and Development, Washington, DC 20460, EPA/600/R-92/128.

Matheron, G. 1963. Principles of geostatistics. Economic Geology, 58(8):1246-1266.

McArthur, RD. 1987. An evaluation of sample designs for estimating a locally concentrated pollutant. Communications in Statistics – Simulation and Computation, 16(3):735-759.

McBratney, AB and Webster, R. 1986. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. Journal of Soil Science, 37:617-639.

McIntyre, GA. 1952. A method for unbiased selective sampling. Journal of Agricultural Research, 3:385-390.

Mesman, M, Rutgers, M and Jensen, J. 2006. Using the triad in site-specific assessment of contaminated soil. *In*: Jensen, J and Mesman, M (Eds). Ecological risk assessment of contaminated land: Decision support for site specific investigations. RIVM report number 711701047, ISBN 978-90-6960-138-0, pp. 41-53.

Monenco (Monenco Ontario Limited). 1981. Evaluation of the potential of peat in Ontario. Occasional Paper No. 7. Prepared for: Ontario Ministry of Natural Resources, Energy Ontario, Ministry of Northern Affairs and Ministry of Energy.

Mood, AM, Graybill, FA and Boes, DC. 1974. Introduction to the theory of statistics, 3rd Edition. McGraw Hill, Toronto.

Moody, M. 2004. Development of terrestrial plant test methods using boreal forest species. Saskatchewan Research Council, Environment and Minerals Division, Saskatoon, SK. Prepared for: Biological Methods Division, Environmental Technology Centre, Environment Canada, Ottawa, ON, pp. 53, Appendices.

Moody, M. 2006. Further development of boreal forest plant toxicity test. Saskatchewan Research Council, Environment and Minerals Division, Saskatoon, SK. Prepared for: Biological Methods Division, Environmental Technology Centre, Environment Canada, Ottawa, ON, pp. 56, Appendices.

Morgan, E and Knacker, T. 1994. The role of laboratory terrestrial model ecosystems in the testing of potentially harmful substances. Ecotoxicology, 3:213-233.

Moser, T, Römbke, J, Schallnass, H-J and Van Gestel, CAM. 2007. The use of the multivariate Principal Response Curve (PRC) for community-level analysis: a case study on the effects of carbendazim on enchytraeids in Terrestrial Model Ecosystems (TME). Ecotoxicology, 16:573-583.

Mueller, DC, Bonner, JS, McDonald, SJ, Autenrieth, RL, Donnelly, KC, Lee, K, Doe, K and Anderson, J. 2003. The use of toxicity bioassays to monitor the recovery of oiled wetland sediments. Environmental Toxicology and Chemistry, 22(9):1945-1955.

Nature Conservancy of Canada. 2009. Alvars of Ontario. The Nature Conservancy of Canada http://www.natureconservancy.ca/site/News2?abbr=on_ncc_&page=newsarticle&id=5961

Neyman, J. 1934. On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. Journal of the Royal Statistical Society. 97:558-606.

New Brunswick Department of Natural Resources, Ecological Classification Working Group. 2007. Our landscape heritage: The story of ecological land classification in New Brunswick. 2nd edition, ISBN 978-1-55396-203-8. http://www.gnb.ca/0399/OurLandscapeHeritage/

Nikolakis, A, Scholz-Starke, B, Heimback, F, Leicher, T, Schaeffer, A and Ross-Nickoll, M. 2007. Terrestrial model ecosystems (TME): A higher-tier semi-field test system to study effects of contaminants on the soil mesofauna. Presented at SETAC North America, 28th Annual Meeting, November 11-15, 2007, Milwaukee, WI, USA.

NJDEP (New Jersey Department of Environmental Protection). 2005. Field Sampling Procedures Manual.

Nöether, GE. 1987. Sample size determination for some common nonparametric tests. Journal of the American Statistical Association, 82(398): 645-647.

NRC (National Research Council Canada). 1995. Toxicity testing to soil ecotoxicology. Workshop Proceedings, November 28-29, 1995 Montreal, Quebec. Hosted by NRC's Biotechnology Research Institute, Environment Canada and the Quebec Ministry of Environment and Wildlife, pp. 35.

NRCan (Natural Resources Canada). 2009a. Surficial materials. Atlas of Canada. http://atlas.nrcan.gc.ca/site/english/maps/environment/land/surficialmaterials/1

NRCan (Natural Resources Canada). 2009b. Wetland diversity. The Atlas of Canada, Natural Resources Canada, Her Majesty the Queen in Right of Canada. http://atlas.nrcan.gc.ca/site/english/maps/environment/ecology/components/wetlanddiversity

OECD (Organisation for Economic Co-operation and Development). 2008. Guideline for testing of chemicals. Predatory mite reproduction test in soil (*Hypoaspis* (Geolaelaps) *aculeifer* Canestrini (Acari: Laelapidae)). No. 226. Paris, France.

Olea, RA. 1984. Sampling design optimization for spatial functions. Journal of the International Association for Mathematical Geology. 16: 369-392.

Olde Venterink, H., Davidson, TE, Kiehl, K. and Lenardson, L. 2002. Impact of drying and re-wetting on N, P and K dynamics in a wetland soil. Plant and Soil, 243: 119-130.

Paetz, A and Wilke, B-M. 2005. Soil sampling and storage. In: Manual for soil analysis – monitoring and assessing soil bioremediation. Margesin, R and Schinner, F (Eds). Springer, Berlin, p. 1-45.

Page-Dumroese, D and Brown, RF. 1999. Comparison of methods for determining bulk densities of rocky forest soils. Soil Science Society of America Journal, 63: 379-383.

Parent, LE and Caron, J. 1993. Physical properties of organic soils. *In*: Carter, M.R. (ed.) Soil Sampling and Methods of Analysis. Canadian Society of Soil Science, CRC Press Taylor and Francis Group, Boca Raton, FL, pp. 441-458.

Pascoe, GA, Blanchet, RJ and Linder, G. 1993. Ecological risk assessment of a metals-contaminated wetland: reducing uncertainty. The Science of the Total Environment, (Supplement 1993):1715-1727.

Pennock, D, Yates, T and Braidek, J. 2008. Soil sampling designs. *In*: Carter, MR and Gregorich, EG (Eds). 2008. Soil sampling and methods of analysis, 2nd Edition. Canadian Society of Soil Science, CRC Press Taylor and Francis Group, Boca Raton, FL, pp. 1-14.

Pitard, FF. 1993. Pierre Gy's sampling theory and sampling practice, 2nd Edition. CRC Press, Boca Raton, FL.

Posthuma, L, Suter GW and Traas, TP. 2002. Species sensitivity distributions in ecotoxicology. CRC Press, Boca Raton.

Prévost, D and Antoun, H. 2008. Root nodule bacteria and symbiotic nitrogen fixation. *In*: Carter, MR and Gregorich, EG (Eds). 2008. Soil sampling and methods of analysis, 2nd Edition. Canadian Society of Soil Science, CRC Press Taylor and Francis Group, Boca Raton, FL, pp. 379-397.

Prietzel, J, Tyufekchieva, N, Eusterhues, K, Kogel-Knabner, I, Thieme, J, Paterson, D, McNulty, I, de Jonge, M, Eichert, D and Salome, M. 2009. Anoxic versus oxic sample pretreatment: Effects on the

speciation of sulphur and iron in well-aerated and wetland soils as assessed by X-ray absorption near-edge spectroscopy (XANES). Geoderma, 153:318-330.

Princz JI, Behan-Pelletier VM, Scroggins RP, and Siciliano, SD. 2010. Oribatid mites in soil toxicity testing – the use of *Oppia nitens* (C.L. Koch) as a new test species. Environmental Toxicology and Chemistry, 29(4):971-979.

Prodromou, KP and Pavlatou-Ve, AS. 2006. Changes in soil pH due to the storage of soils. Soil Use and Management, 14(3):182-183.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, ISBN 3-900051-07-0 http://www.R-project.org

Rahn, J. 2008. Measurement endpoints for soil microbial toxicity. Internal report, Biological Methods Division, January, 2008, Ottawa, Ontario.

Reinhardt, CH, Cole, CA and Stover, LR. 2000. A method for coring inland, freshwater wetland soils. Wetlands, 20(2):422-426

Renoux, AY, Tyagi, RD and Samson, R. 1999. Effects of irradiation and freezing on toxicity of sewage sludge elutriate samples. Water Quality Research Journal of Canada, 34(4):589-597.

Ripley, BD. 1981. Spatial Statistics. Wiley, New York.

Römbke, J. 2006. Tools and techniques for the assessment of ecotoxicological impacts of contaminants in the terrestrial environment. Human and Ecological Risk Assessment: An International Journal, 12(1):84-101.

Römbke, J and Amorim, M. 2004. Tackling the heterogeneity of soils in ecotoxicological testing: An EURO-Soil based approach. Journal of Soils and Sediments, 4:276-281.

Römbke, J, Heimbach, F, Hoy, S, Kula, C, Scott-Fordsmand, J, Sousa, J, Stephenson, G and Weeks, J. 2003. Effects of plant protection products on functional endpoints in soil. Workshop Proceedings, EPFES, Lisboa 2002. Society of Environmental Toxicology and Chemistry, 2003, Pensacola, Florida, p. 92.

Römbke, J, Eisenträger, A, Hund-Rinke, K, Jänsch, S, Neumann-Hensel, H and Schallnaß, H-J (Eds). 2006a. Erprobung and Vorbereitung einer praktischen Nutzung ökotoxikologischer Testsysteme. Gefördert vom BMBF (PT Julich), ERNTE, September 2006, Sidus Verlag. ISBN 3-937451-10-2

Römbke, J, Jänsch, S and Scroggins, R. 2006b. Identification of potential organisms of relevance to Canadian boreal forest and northern lands fro testing of contaminated soils. Environmental Reviews, 14:137-167.

Sauvé, S. 1997. Speciation of metals in soils. *In*: Bioavailability of metals in terrestrial ecosystems: Importance of partitioning for bioavailability to invertebrates, microorganisms and plants. SETAC, Pensacola, FL, pp 7-37.

Schaeffer, A, Van den Brink, P, Heimbach, F, Hoy, S, De Jong, F, Römbke, J, Sousa, JP and Ross-Nickoll, M. 2008. Semi-field methods are a useful tool for the environmental risk assessment of pesticides in soil. Environmental Science and Pollution Research, 15:176-177.

Schoeneberger, PJ, Wysocki, DA, Benham, EC and Broderson, WD (Eds). 2002. Field book for describing and sampling soils. Version 2.0. Natural Resources Conservation Service, National Soil Survey Center, Lincoln NE.

Sheppard, SC and Addison, JA. 2008. Soil sample handling and storage. *In*: Carter, MR and Gregorich, EG (Eds). 2008. Soil sampling and methods of analysis, 2nd Edition. Canadian Society of Soil Science, CRC Press Taylor and Francis Group, Boca Raton, FL, pp. 39-49.

Sheppard, MI, Tarnocai, C and Thibault, DH. 1993. Sampling organic soils. *In*: Soil Sampling and Methods of Analysis, Carter, MR (Ed), Canadian Society of Soil Science, CRC Press Taylor and Francis Group, Boca Raton, FL, pp. 423-439.

Sheremata, TW, Raymond, NY and Guiot, SR. 1997. Simulation and sterilization of a surrogate soil organic matter for the study of the fate of trichloroethylene in soil. Communications in Soil Science and Plant Analysis, 28(13&14):1177-1190.

Sims, RA, Towill, WD, Baldwin KA and Wickware, GM. 1997. Field Guide to the forest ecosystem classification for northwestern Ontario, 2nd ed. Field Guide FG-03. Forestry Canada and Ontario Ministry

of Natural Resources, Thunder Bay, Ontario, pp. 176.

Singer, DA. 1972. Elipgard, a Fortran IV program for calculating the probability of success in locating elliptical targets with square, rectangular and hexagonal grids. Geocomm Programs, 4:1-16.

Slattery, WJ and Burnett, VP. 1992. Changes in soil pH due to long term soil storage. Australian Journal of Soil Research, 30(2):169-175.

Sokal, RR and Rohlf, FJ. 1995. Biometry, 3rd Ed. W.H. Freeman & Co., New York.

Soil Landscapes of Canada Working Group. 2007. Soil landscapes of Canada v3.1.1 Agriculture and Agri-Food Canada. (digital map and database at 1:1 million scale).

Soil Survey Division Staff. 1993. Soil survey manual. Soil Conservation Service. U.S. Department of Agriculture Handbook 18.

Spurgeon, DJ, Svendson, C, Hankard, PK, Toal, M, McLennan, D, Wright, J, Walker, L, Ainsworth, G, Wienberg, C and Fishwick, SK. 2004. Application of sublethal ecotoxicological tests for measuring harm in terrestrial ecosystems. Environment Agency, Bristol, UK, Environment Agency Science Report P5-063/TR2, pp. 103.

Spurgeon, D, Svendson, C, Hankard, P, Weeks, J, Kille, P and Fishwick, S. 2002. Review of sublethal ecotoxicological tests for measuring harm in terrestrial ecosystems. Environment Agency, Bristol, UK, Environment Agency Science Report P5-063/TR1, pp. 191.

Stantec (Stantec Consulting Ltd.). 2004. Framework foundation and guidance for Tier 2 site-specific development of soil contact standards for PHC-contaminated sites: Literature review. Prepared for: Petroleum Technology Alliance Canada, Calgary, AB and Canadian Petroleum Products Institute, Ontario Division, North York ON, pp.83, Appendix.

Stantec (Stantec Consulting Ltd.). 2005. Ecotoxicity assessment of site soils from the Swan Hills area contaminated predominantly with F3. Prepared for: Komex International Ltd., Calgary, AB pp. 32, Appendices.

Stantec (Stantec Consulting Ltd.). 2006. Summary of chronic screening toxicity test results with reference soil and soils contaminated with petroleum hydrocarbon residuals. Prepared for: Petroleum Technology Alliance Canada, Calgary, AB pp. 38, Appendices.

Stantec (Stantec Consulting Ltd.). 2008. Earthworm chronic testing Phase 2: Identifying effective soil conditioners and amendment levels in clay and sand reference soils for use in toxicity testing. Prepared for: Petroleum Technology Alliance Canada, Calgary, AB pp.42, Appendices.

Steevens, JA, Vansal, SS, Kallies, KW, Knight, SS, Cooper, CM and Benson, WH. 1998. Toxicological evaluation of constructed wetland habitat sediments utilizing *Hyalella azteca* 10-day sediment toxicity test and bacterial bioluminescence. Chemosphere, 36(15):3167-3180.

Stevens, DL and Jensen, SF. 2007. Sample design, executions and analysis for wetland assessment. Wetlands, 27(3):515-523

Stephenson, GL, Kuperman, RK, Linder, GL and Visser, S. 2002. Toxicity tests for assessing contaminated soils and ground water. In: Environmental Analysis of Contaminated Sites. Sunahara, GL, Renoux, AY, Thellen, C, Gaudet, CI and Pilon, A (Eds). 2002. John Wiley and Sons Ltd., p. 25-44.

Stephenson, GL. 2003. Terrestrial test methods for plants and soil invertebrates. Ph.D. thesis, Department of Environmental Biology, University of Guelph, Guelph, ON, pp. 230, Appendices.

Stephenson, G, Feisthauer, N, Bessie, K and Scroggins, R. 2008. Terrestrial toxicity testing in support of site-specific risk assessments: An integrative tool for contaminated site management. Proceedings of the Federal Contaminated Sites National Workshop, Vancouver, BC, April 28-May 2, 2008, hosted by the Real Property Institute of Canada www.rpic-ibic.ca

Strandberg, B, Axelson, JA, Pedersen, MB, Jensen, J and Attrill, MJ. 2006. Effect of a copper gradient on plant community structure. Environmental Toxicology and Chemistry, 25(3):743-753.

Swift, MJ, Heal, OW and Anderson, JM. 1979. Decomposition in terrestrial ecosystems. Blackwell Scientific Publications, Oxford.

Tarnocai, C. 1993. Sampling frozen soils. *In*: Soil Sampling and Methods of Analysis, Carter, MR (Ed), Canadian Society of Soil Science, CRC Press

Taylor and Francis Group, Boca Raton, FL, pp. 755-765.

Tarnocai, C. 2009. Arctic permafrost. *In*: Permafrost Soils, Soil Biology 6. Margesin, R (Ed), Springer-Verlag, Berlin, pp. 3-16.

Thomas, JM, McKenzie, DH and Eberhardt, LL. 1981. Some limitations of biological monitoring. Environment International, 5:3-10.

Thomas, JM, Skalski, JR, Cline, JF and McShane, MC. 1986. Characterization of chemical waste site contamination and determination of its extent using bioassays. Environmental Toxicology and Chemistry, 5:487-501.

Thompson, SK. 1992. Sampling. John Wiley & Sons, Toronto.

Thompson, KC, Wadhia, K and Loibner, AP (Eds). 2005. Environmental toxicity testing. Blackwell Publishing, CRC Press.

Tiessen, H and Moir, JO. 1993. Total and organic carbon. Chapter 21 *In*: Carter, MR (Ed), 1993. Soil Sampling and Methods of Analysis. Canadian Society of Soil Science, CRC Press Taylor and Francis Group, Boca Raton, FL, pp. 187-199.

Topp, E, Tien, Y-C and Hartmann, A. 2008. Bacterial community analyses by denaturing gradient gel electrophoresis. *In*: Carter, MR and Gregorich, EG (Eds). 2008. Soil sampling and methods of analysis, 2nd Edition. Canadian Society of Soil Science, CRC Press Taylor and Francis Group, Boca Raton, FL, pp. 567-576.

Trevors, JT. 1996. Sterilization and inhibition of microbial activity in soil. Journal of Microbiological Methods, 26:53-59.

UMA (UMA Engineering Ltd.). 2008. Risk based salinity guidelines for peatlands Phase 1: Literature review and gaps analysis. Prepared for the Canadian Association of Petroleum Producers Peatland Salinity Working Group, Calgary, AB, pp. 72, Appendices.

University of Alberta. Alberta Institute of Pedology. 1975. Reconnaissance soil survey of the Sand River area. 73L. University of Alberta Bulletin N. SS 15. Alberta Institute of Pedology Report No. S-74-34. Compiled by: Kocaoglu, S.S. Agriculture Canada. Soil Research Institute.

University of Saskatchewan. 2009. College of Agriculture and Bioresources. Department of Soil Science. http://soilsofcanada.ca/orders/index.php.

USEPA (United States Environmental Protection Agency). 1986. SW-846 Test methods for evaluating solid waste, physical/chemical methods. Chapter Nine: Sampling Plan, pp 79. http://www.epa.gov/epawaste/hazard/testmethods/sw846/online/index.htm.

USEPA (US Environmental Protection Agency). 1989. Methods for Evaluating the Attainment of Cleanup Standards, Volume 1: Soils and Solid Media (PDF) (EPA 230/02-89-042, February 1989).

USEPA (United States Environmental Protection Agency). 1993. Guide for conducting treatability studies under CERCLA. Biodegradation remedy selection, interim guidance. EPA/540/R-93/519a. Superfund Innovative Technology Evaluation, Office of Solid Waste and Emergency Response, Washington, DC, 20460.

USEPA (United States Environmental Protection Agency). 2002a. Guidance on choosing a sampling design for environmental data collection for use in developing a quality assurance project plan. EPA QA/G-5S. Office of Environmental Information, Washington, DC, 20460, EPA/240/R-02/005, December 2002.

USEPA (United States Environmental Protection Agency). 2002b. RCRA Waste Sampling Draft Technical Guidance Planning, Implementation and Assessment. Office of Solid Waste. EPA530-D-02-002.

USEPA (United States Environmental Protection Agency). 2002c. Methods for the determination of total organic carbon (TOC) in soils and sediments. Brian A. Schumacher, USEPA, Environmental Sciences Division National Exposure Research Laboratory. Ecological Risk Assessment Support Center. NCEA-C-1282 April 2002.

USEPA (United States Environmental Protection Agency). 2006. Sample collection procedures for radiochemical analytes in environmental matrices. Office of Research and Development, National Homeland Security Research Center, Cincinnati, Ohio, 45268, EPA/600/S-07/001.

USEPA (United States Environmental Protection Agency). 2007. Framework for metals risk assessment. Office of Research and Development,

National Center for Environmental Assessment, Washington DC. EPA/120/R-07/001.

US Navy, 2009. Navy environmental compliance sampling and field testing procedures manual. NAVSEA T0300-AZ-PRO-010.

Valliant, R, Dorman, AH and Royall, RM. 2000. Finite population sampling and inference – A prediction approach. John Wiley & Sons, Toronto.

Vischetti, C, Corti, G, Monaci, E, Cocco, S, Coppola, L and Agnelli, A. 2010. Pesticide adsorption and degradation in fine earth and rock fragments of two different origin. Geoderma, 154(2010):348-352.

Walmsley, M, Utzig, G, Vold, T, Moon, D, and van Barneveld, J. 1980. Describing ecosystems in the field. RAB Technical Paper 2. Land Management Report No. 7. Province of British Columbia. Ministry of Environment, Lands and Parks, Ministry of Forests. ISSN 0702-9861. http://www.for.gov.bc.ca/hfd/pubs/Docs/Mr/Lmr/Lmr007.pdf

Warren-Hicks, W, Parkhurst, BR and Baker Jr, SS. 1989. Ecological assessment of hazardous waste sites. United States Environmental Protection Agency, Environmental Research Laboratory, Corvallis, OR, 97333, EPA 600/3-89/013.

Weeks, JM, Sorokin, N, Johnson, IJ, Whitehouse, P, Ashton, D, Spurgeon, D., Hankard, P, Svendson, C and Hart, A. 2004. Biological test methods for assessing contaminated land: Stage 2 – a demonstration of the use of a framework for the ecological risk assessment of contaminated land. Environment Agency, Bristol, UK, Environment Agency Science Report P5-069/TR1, pp. 106.

Weider, RK and Lang, GE. 1986. Fe, Al, Mn and S chemistry of Sphagnum peat in four peat lands with different metal and sulphur input. Water, Air and Soil Pollution, 29: 309-320.

Weinfurtner, K and Kördel, W. 2007. Umweltprobenbank des Bundes. Guidelines for sampling and sample processing. Soil. Guidelines for sampling, transport, storage and chemical characterization of environmental and human-organ samples, Umweltbundesamt, Germany, August 2007 Draft.

Wendling, LA, Kirby, JK and McLaughlin, MJ. 2009. Aging effects on cobalt availability in soils. Environmental Toxicology and Chemistry, 28(8):1609-1617.

Weyers, A, Sokull-Klüttegn, B, Knacker, T, Martin, S and Van Gestel, CAM. 2004. Use of Terrestrial Model Ecosystem data in environmental risk assessment for industrial chemicals, biocides and plant protection products in the EU. Ecotoxicology, 13:163-176.

Wiles, JA and Krogh, PH. 1998. Tests with the collembolans *Isotoma viridis*, *Folsomia candida*, and *Folsomia fimetaria*. *In*: Handbook of Soil Invertebrate Toxicity Tests. Løkke, H and van Gestel, CAM, eds., John Wiley, London, United Kingdom, p. 131-156.

Yfantis, EA, Flatman, GT and Behar, JV. 1987. Efficiency of kriging estimation for square, triangular, and hexagonal grids. Mathematical Geology, 19:183-205.

Yukon Environment. 2002. Protocol for the Contaminated Sites Regulation under the Environment Act, Protocol No. 9: Determining background soil concentration. http://www.environmentyukon.gov.yk.ca/pdf/protocol9.pdf

Zagury, GJ, Dudal, Y, Bureau, J, Bastien, C and Chassé, R. 2002. Sample handling and preparation for estimation of mobility, bioavailability and toxicity of contaminants in soils. In: Environmental analysis of contaminated sites. Sunahara, GI, Renoux, AY, Thellen, C, Gaudet, CL and Pilon, A (Eds.). Ecological and Environmental Toxicology Series, John Wiley and Sons, Ltd., West Sussex, pp. 9-23.

Zar, JH. 1999. Biostatistical analysis, 4th Ed. Prentice-Hall, Toronto.

Zirschky, J and Gilbert, RO. 1984. Detecting hot spots at hazardous waste sites. Chemical Engineering, 91:97-100.

Zoltai, SC. 1978. A portable sampler for perennially frozen stone-free soils. Canadian Journal of Soil Science, 58:521-523.

Zoltai, SC and Vitt, DH. 1995. Canadian wetlands: Environmental gradients and classification. Plant Ecology, 118:131-137.

Appendix A

Case Studies of the Application of Biological Testing

Section 2 provides recommendations for the use of biological tests in assessing the toxicity of contaminated land, and brief descriptions of each test. Appendix A provides four examples of how biological tests were used as tools for the assessment of contaminated lands. Three of the four are Canadian examples; the fourth is international.

A.1 Case Study #1: The use of biological toxicity testing as a tool in a site-specific risk assessment of biotreated petroleum hydrocarbon-contaminated soils

A preliminary quantitative risk assessment was conducted on biotreated surface soils on a site at a former land treatment facility. The characterization of the surface soils at the site indicated that barium, salinity, and petroleum hydrocarbons (PHCs) CCME Fraction 3 (C_{16}-C_{34}) exceeded acceptable screening benchmarks. This indicated that ecotoxicity data would be required to derive site-specific, risk-based remediation objectives, and to determine the risk to ecological receptors of exposure to soils with these substances via the soil contact pathway. A weight-of-evidence approach was used to characterize risk at the site. Toxicity tests to evaluate the effect of direct contact of organisms to representative site soil samples were conducted and used as one of five lines of evidence in the weight-of-evidence approach. The results of these tests were integrated with the other lines of evidence (calculation of soil quality criteria for barium and the human ingestion exposure pathway; results for a DTPA-extractable barium test; and estimation of the influence of threshold effect concentration values for reduction of vegetation) to characterize the risk of adverse effects to the identified receptors of concern. Uncertainty analysis was undertaken to assess the degree of confidence in the characterization of risk and to determine if a detailed quantitative risk assessment was required.

The toxicity tests were conducted using the draft Environment Canada biological methods under development at that time. The test species included acute (7-d to 14-d) and definitive (14- to 21-d) tests with barley *(Hordeum vulgare* var. Chapais*)*, northern wheatgrass *(Elymus lanceolatus)*, and red clover *(Trifolium pratense)*, and an acute (14-d) and two reproduction (56- and 63-d) tests with the earthworm *Eisenia andrei*. Toxicity was assessed by comparing the performance of the test species in the biotreated soils relative to the performance of the test species in a field-collected site-specific reference (uncontaminated) control soil with physical and chemical characteristics similar to those of the biotreated soil. An artificial soil (AS) and a generic clay loam reference soil (RS) were included for QA/QC purposes in all but the 56-d earthworm reproduction test.

No adverse effects on acute (14-d) or chronic (28- or 35-d) earthworm survival was observed; no adverse effects on earthworm reproduction (number of progeny produced, number of hatched and unhatched cocoons, progeny wet and dry mass) were observed following exposure to the biotreated soil (Figure A.1). Seedling emergence was unaffected following short-term exposure for all three species. Following longer, definitive exposure, seedling emergence for red clover and northern wheatgrass was unaffected; however, barley emergence was reduced by 16% in the biotreated soil. Red clover and northern wheatgrass seedling growth (shoot and root length, wet and dry masses) was not affected in plants grown in the biotreated soil relative to those grown in the site-specific reference soil following acute or definitive exposure. In contrast, some barley growth endpoints were adversely affected following both acute (7-d) and definitive (14-d) exposure to the biotreated soil (Figure A.2). The adverse effects on growth and emergence of barley were inconsistent between the acute and chronic tests and among endpoints.

The ecotoxicity assessment generated 58 toxicity endpoints (6 earthworm survivorship and 10 reproduction endpoints and 6 seedling emergence and 36 growth endpoints). Following the statistical analyses and uncertainty evaluation of these data, it was concluded that there were no observable adverse effects on survival, growth, and reproduction of earthworms exposed for either acute or prolonged periods to the biotreated soils. There was no toxicity associated with acute or prolonged exposure of clover and northern wheatgrass to the biotreated soils; however, barley emergence and growth was adversely affected. The results for barley indicated that the biotreated soil was less than optimal for growth; however, the inconsistency among endpoints suggests that the physicochemical characteristics of the soil influenced the toxicity test results.

The integration of toxicity assessment results with other lines of evidence (literature data, soil leachate test data, measured total and available contaminant concentrations, derivation of a human health site-specific soil quality objective for barium) revealed that although the highest potential for adverse effects from exposure to the biotreated surface soils was through direct soil contact, there was a high degree of confidence that the site soil was not toxic to receptors; and therefore, it represents a negligible risk to potential receptors. The toxicity data from this Tier 2 site-specific ecotoxicity testing, in consideration of other ancillary information, contributed to the regulatory closure for this site.[35]

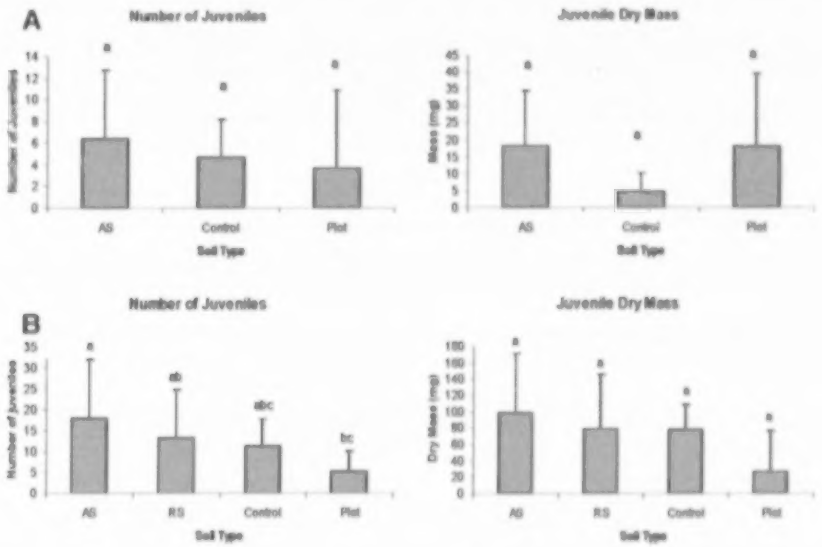


Figure A.1 Progeny production and dry mass of earthworms exposed to two or three negative control soils (AS – artificial soil; RS – in-house reference soil; Control - site-specific reference control soil) and the biotreated surface soil (Plot) for either 56 (A) or 62 days (B). Columns represent mean values, bars indicate standard deviations, and columns with similar letters are not significantly different at $P > 0.05$.

[35] Additional details of this study can be found in Stephenson *et al.*, 2008.

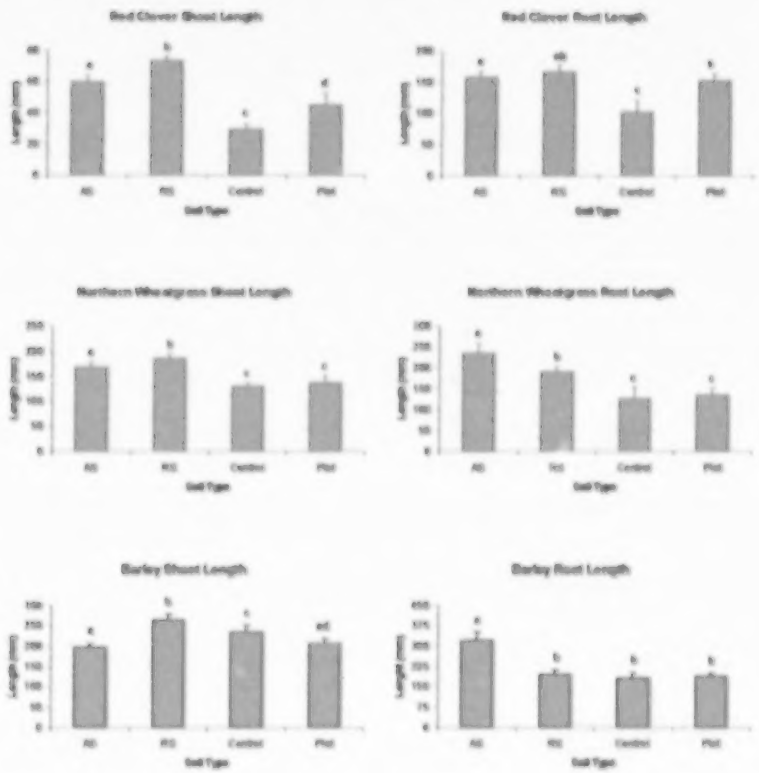


Figure A.2 Shoot and root length of clover, northern wheatgrass, and barley exposed to the three negative control soils (AS – artificial soil; RS – in-house reference soil; Control – site-specific reference control soil) and the on-site biotreated surface soil (Plot) for either 21, 21, and 14 days, respectively. Columns represent mean values, bars indicate standard deviations, and columns with similar letters are not significantly different at $P > 0.05$.

A.2 Case Study #2: Derivation of site-specific remedial objectives using biological testing for a metal-contaminated site

Terrestrial ecotoxicity testing was conducted in support of a site-specific ecological risk assessment of soil contaminated with a mixture of metals and metalloids due to transport operations associated with historical smelting operations. Site-specific ecotoxicity data were generated from a test battery consisting of definitive (14- to 21-d) plant tests with barley, perennial ryegrass *(Lolium perenne)*, red clover, red fescue *(Festuca rubra)* and reproduction (35- to 63-d) tests with the earthworm *Eisenia andrei* and the collembola *Orthonychiurus folsomi* using the Environment Canada test methods (EC, 2004a, 2005a, 2007a). Contamination of the soil was primarily limited to the top few centimetres of the surface; therefore, surficial soil was collected for testing (F, H, and uppermost A horizons only). Because the metal contamination was associated with the organic surface layer, if the soils that were prepared for testing included any thatch or plant roots that were sieved from the soil samples (6 mm-mesh), these were ground and incorporated into the soil test samples.

The objective of the toxicity testing was not to determine the average toxicity associated with the site, but rather the relationship among contaminant concentrations, contaminant mixtures, and soil physicochemical characteristics because all three of these factors, which were widely distributed in micro sites throughout the site, have the potential to influence toxicity. Because the site was very large and heterogeneous, it was not cost-effective to test multiple replicates of soil for each stratum of the site. Instead, a regression-based sampling and testing design was used in which each microsite of interest was represented by one soil sample. Many soil samples were collected and each one was tested as one laboratory replicate per species. The current standard procedure at the time was to test multiple laboratory subsamples per species for each soil sample collected from the field.

Because the soil samples collected represented contaminant concentration-gradients (for mixtures of contaminants), multiple regression models could be created for most of the toxicity endpoints for each of the six test species. These models were developed so they could be used to predict the soil contaminant concentrations that would cause any user-defined percent effect of test species performance relative to performance in the control soil [the concentration of contamination that would cause, for example, a 25% decrease in species performance relative to their performance in a similar reference (uncontaminated) soil]. The acceptable percent threshold effect used was set by the regulatory agency under whose jurisdiction the management of the contaminated land fell. Because soil characteristics such as pH, texture, fertility, and cation exchange capacity varied spatially, and because these characteristics have the potential to influence the toxicity test results, the modelling of the toxicity data incorporated the possible influence of these and other soil characteristics (the soil characteristics were treated as separate variables). The regression models adjusted for the soil characteristics and using the percent effect thresholds set by the regulatory agency were then used as predictive tools to estimate site-specific remedial objectives for the protection of ecological receptors.

A.3 Case Study #3: Contribution of site-specific ecotoxicity testing to achieve site closure of petroleum hydrocarbon contaminated land

Alberta Environment has recently published Alberta Tier 1 and Tier 2 soil and groundwater remediation guidelines (AE, 2007a, 2007c). Tier 1 guidelines for petroleum hydrocarbons (PHCs) in soil may be overly conservative for sites where PHCs have reduced bioavailability and/or toxicity to soil organisms because they are highly weathered or aged. The development of Tier 2 site-specific objectives for direct exposure pathways (ecological soil contact) can be developed for PHCs when conducting a site-specific risk assessment (AE, 2007a). Guidance for the use of the site-specific ecotoxicity data is provided in the draft Tier 2 Eco-contact Guideline Derivation Protocol (AE, 2007b). In order to achieve regulatory closure for a site, proponents must subject site soils to a Tier 2 Pass/Fail ecotoxicity program. If soils are deemed non-toxic to the minimum required test battery (if the soils "pass") according to criteria in the protocol, then regulatory closure might be achieved.

A case study is described in which ecotoxicity data were generated from soils contaminated with weathered PHCs and these data then subjected to the Tier 2 Pass/Fail criteria. Soil from this site was contaminated with weathered PHCs and had undergone bioremediation, but residual CCME PHC Fraction 3 concentrations still exceeded Tier 1 levels. Regulatory closure on this site for natural areas, agriculture, and residential/parkland land use was sought, in part, through the use of the Tier 2 Pass/Fail criteria. In this draft guidance, a Tier 2 pass for natural areas, or for agriculture and residential/parkland uses must meet the following criteria:

- when statistically significant differences are identified between reference and contaminated soils, the differences must be $\leq 25\%$ for at least 75% of the endpoints;
- test organism mortality must be no greater in the contaminated soil than in the reference soil;
- invertebrate reproduction in the contaminated soil must not be $< 50\%$ of that in the reference soil;

- no more than one endpoint per test species may exceed a 25% difference between contaminated and reference soil; and,
- the experimental design must have adequate power to detect a difference of 25% or more between treatments.

Representative composite contaminated site and reference soil samples were collected for testing. Sample handling was restricted to homogenization at the laboratory. The performance of the test organisms in the contaminated site soil was compared to the performance of test organisms in a field-collected reference control soil with physical and chemical characteristics similar to those of the contaminated site soils but free of contamination. The test battery minimum was met and included definitive (21-d) tests with alfalfa *(Medicago sativa)* and northern wheatgrass, and reproduction tests with *Eisenia andrei* (63-d) and the collembola *Folsomia candida* (28-d). As required, Environment Canada standard methods were used (EC, 2004a, 2005a, 2007a).

Table A.1 provides a summary of the magnitude of the difference of all plant and invertebrate endpoints between the contaminated soil treatments and the reference control soil, and indicates whether or not the difference was statistically significant.

The experimental design of the tests had sufficient power to detect a difference of 25% or more among soil treatments. Only 3 of 16 endpoints in the contaminated soil had statistically lower values relative to the same endpoints in the reference soil. Of these 3 endpoints, only 2 endpoints (northern wheatgrass root dry mass and number of earthworm progeny produced) had a difference > 25%. Earthworm reproduction in the contaminated soil was < 50% (42%) of that in the contaminated soil, which violates one of the Tier 2 Pass criteria.

Earthworm reproduction is very sensitive to organic matter content level in soils, particularly for *Eisenia* species, and the threshold levels for optimal earthworm reproduction are generally between 3% and 4% (Jänsch *et al.*, 2005). The organic matter content in the contaminated soil was more than 2.5-fold less than that in the reference soil and was below the 3% to 4% threshold level. Therefore, it is likely that the decreased reproduction in the impacted soils was caused by a combination of the presence of PHCs and the lower organic matter content in the contaminated soil relative to the reference soil.

Using the criteria described by the draft Tier 2 Protocol, the results of this ecotoxicity assessment would fail the Tier 2 Pass/Fail test for all land uses due to the 58% reduction in earthworm reproduction in the contaminated soil relative to earthworm reproduction in the reference soil. However, the low organic matter content in the contaminated soil relative to the reference soil likely adversely effected earthworm reproduction independently of, or in combination with, the effect of the PHCs. As a result, and in consideration of ancillary information, the toxicity data from this Tier 2 site-specific ecotoxicity testing contributed to the regulatory closure for this site.

Table A.1 Results of the toxicity testing program from Case Study #3 expressed as the percent difference from the reference soil for each measurement endpoint for each test and test species.*

Toxicity test and measurement endpoints	Percent difference from reference control soil test performance PHC-contaminated Soil
Alfalfa	
Emergence	-5
Shoot length	**-10**
Root length	16
Shoot dry mass	-8
Root dry mass	11
Northern Wheatgrass	
Emergence	0
Shoot length	-4
Root length	**24**
Shoot dry mass	2
Root dry mass	**41**
Eisenia andrei	
Adult 35-d survival	10
Number of juveniles produced	**58**
Juvenile wet mass	17
Juvenile dry mass	17
Folsomia candida	
Adult 28-d survival	-25
Number of juveniles produced	-66
Number of endpoints	16
Number of endpoints adversely affected**	2

*Values in bold are significantly different from the reference soil. Values that are negative indicate that the measurement was greater than in the reference soil.
**According to the Tier 2 Pass/Fail criteria

A.4 Case Study #4: Use of site-specific toxicity testing in contaminated land management in Germany

Under the German Federal Soil Protection Act and the supplemental German Federal Soil Protection Ordinance contaminated soil must be assessed for its ability to provide suitable (e.g., non-toxic) habitat for soil-dwelling organisms (a "habitat function" criterion) and for its capacity to prevent the leaching of soil contaminants to groundwater (a "retention function" criterion). Although these soil and groundwater protection goals have been legislated since 1999, there is still a lack of clear guidance on how to meet both of these objectives, outside of the provision of a few chemical criteria for selected pollutants. A collaborative research project among government, academic, and private research laboratories was initiated

in 2002 by the German federal government to develop a practical, broadly applicable, and scientifically defensible framework to meet these objectives (Römbke *et al.*, 2006a).

The framework consists of the application of a battery of site-specific terrestrial and aquatic toxicity tests to contaminated soil. Soil toxicity tests are used to evaluate the quality of contaminated soil as habitat and include four tests with short durations (6-h to 5-d) that are used for on-site screening, and three chronic tests conducted off-site in a laboratory that evaluate sensitive endpoints such as growth and reproduction. All tests can be applied before or during site remediation. The on-site screening tests include: two functional tests with indigenous microorganisms [microbial respiration (ISO, 2002e) and nitrification (ISO, 2004b)]; one bacterial *(Arthrobacter globiformis)* soil-contact toxicity test (ISO, 2008b) and a shortened (24-h) earthworm *(Eisenia andrei)* avoidance test (ISO, 2008a). The three chronic tests include a 56-d earthworm reproduction test (ISO, 1998), a 28-d collembola reproduction test (ISO, 1999) and a 35-d life-cycle plant test (ISO, 2005f). The test species are *Eisenia andrei*, *Folsomia candida,* and *Brassica rapa* CrGC for the chronic earthworm, collembolan, and plant test, respectively.

The potential for contamination of groundwater from soil leachate is determined by evaluating the aquatic toxicity of the soil elutriate. These tests are limited to three on-site rapid (6- to 72-h) screening tests: a bacterial *(Vibrio fischeri)* luminescence inhibition assay (ISO, 2007b); an umuC genotoxicity assay with *Salmonella choleraesius* subsp. *choleraesius* (ISO, 2000); and, an algal *(Desmodesmus subspicatus)* growth inhibition test (ISO, 2004c). These aquatic tests were not included in Table 2 (Section 2) because they are tests on soil elutriate, not whole soil, and as such are outside the scope of this guidance document. They are described in this case study to provide a more comprehensive description of the project.

This framework was validated in the assessment of two contaminated sites in Germany, both in the city of Hamburg. One site, Grasbrook, is a historical gas works with soil contaminated with elevated (up to 4600 mg/kg) concentrations of polycyclic aromatic hydrocarbons (PAHs). The other site, Schlachthofstrasse, is a former industrial landfill contaminated with heavy metals and PAHs.

At the Grasbrook site, soils were excavated and stockpiled nearby according to PAH concentration. The soil concentrations ranged from 10 to 1100 mg PAH/kg soil. The framework was applied to soil samples collected from the stockpiles and based on the results, were classified as likely having low risk, indeterminate risk (requiring a "case-by-case" investigation), or high risk to soil organisms, based on the toxicity and chemistry data. Of the 12 samples tested, direct soil toxicity was confirmed in 4 of the 5 samples that were considered to constitute a high risk based on the chemistry data, and toxicity was found in 3 of the 6 samples that were considered to have low impact based on the chemistry data. Only 1 sample was "indeterminate" based on chemistry data but 5 samples were "indeterminate" based on the toxicity data. Relatively low (33%) agreement was found between the chemistry and toxicity data. These results highlight the value of assessing contaminated land based on at least two lines of evidence (toxicity and chemistry) using a weight-of-evidence approach to prevent erroneous management decisions and/or to enable management decisions to be made with more confidence.

In contrast, at the Schlachthofstrasse site there was 75% agreement between the chemistry and toxicity data for the four soil samples evaluated. Three of the four samples were considered to have a high risk of impact and the other sample was considered to have "indeterminate" risk. From the results from both sites, it seems that conclusions of high risk based on chemistry data are more predictive of toxicity than conclusions of intermediate or low risk. The results from both case studies demonstrate that the proposed framework is a sensitive approach and is practical to apply. The use of this site-specific toxicity testing framework resulted in cost savings because it reduced the volume of hazardous material requiring disposal and allowed other uses for the marginally contaminated soils.

Appendix B

Ecological Classification Resources

B.1 National Ecological Land Classification Systems

The National Ecological Framework for Canada defines the ecological land classification and the ecozones and ecoregions of Canada, the reference for this document is:

- Marshall, IB and Schut, PH. 1999. A National Ecological Framework for Canada. Overview. Ecosystems Science Directorate, Environment Canada and Research Branch, Agriculture and Agri-Food Canada. http://sis.agr.gc.ca/cansis/nsdb/ecostrat/intro.html

The Canadian Forest Service (Natural Resources Canada) currently has a Canadian Forest Ecosystem Classification project. The objective of this project is to correlate provincial and territorial classifications into a common national system. This forest classification system will correspond to the International Classification of Ecological Communities in Canada and the United States.

- Study Leader: Mr. Ken Baldwin, Forest ecologist (Great Lakes Forestry Centre, Canadian Forest Service, Natural Resources Canada. Ken.Baldwin@NRCan-RNCan.gc.ca (705) 541-5642.

B.2 Regional Ecological Land Classification Systems

Regional classification systems have also been or are being developed; these are typically on a provincial and/or territorial level. Depending on the province or territory, regional classification systems may be based on forest regions, natural resource management, land use or conservation planning, and/or they may be described by field guides or biodiversity monitoring programs. Table B.1 lists Canadian ecological land classification systems available by region.

Table B.1 Ecological land classification systems available by region in Canada.

Province/ Territory/ Region	Type	Ecological land classification system/reference
British Columbia	Forest-based regional classification system	BC Ministry of Forests and BC Ministry of Environment, Lands and Parks. 1998. Field Manual for Describing Terrestrial Ecosystems. Land Management Handbook Number 25. ISSN 0229-1622. (http://www.for.gov.bc.ca/hfd/pubs/Docs/Lmh/Lmh25/Lmh25.pdf) This is the third edition of this reference; references for the first two editions (provided below) are also used: • 2nd edition: Luttmerding, HA, Demarchi, DA, Lea, EC, Meidinger DV, and Vold T (eds). 1990. Describing ecosystems in the field. Second edition. MOE Manual 11. Province of British Columbia. Ministry of Environment,

		Lands and Parks, Ministry of Forests. ISSN 0821-0640. (http://www.for.gov.bc.ca/hfd/pubs/docs/mr/mr074.pdf)
		• 1st edition: The "cheat sheets" for classification are only contained within the first edition, which is Walmsley M, Utzig G, Vold T, Moon D, van Barneveld J. 1980. Describing ecosystems in the field. RAB Technical Paper 2. Land Management Report No. 7. Province of British Columbia. Ministry of Environment, Lands and Parks, Ministry of Forests. ISSN 0702-9861. (http://www.for.gov.bc.ca/hfd/pubs/Docs/Mr/Lmr/Lmr007.pdf)
		Province of British Columbia. 2007. Biogeoclimatic ecosystem classification program. Forest Service of British Columbia, Research Branch. (http://www.for.gov.bc.ca/hre/becweb/resources/classificationreports/index.html)
	Natural resource management	Demarchi, DA. 1996. An introduction to the ecoregions of British Columbia.Wildlife Branch, Ministry of Environment, Lands and Parks, Victoria, BC. (http://www.env.gov.bc.ca/ecology/ecoregions/intro.html)
Yukon	Field guide	Smith CAS, Meikle JC, Roots CF (eds). 2004. Ecoregions of the Yukon landscapes. Agriculture and Agri-Food Canada, PARC Technical Bulletin No. 04-01, Summerland, BC.
	Regional land use planning/ conservation	Meikle J, Waterreus M. 2008. Ecosystems of the Peel watershed: A predictive approach to regional ecosystem mapping. Yukon Fish and Wildlife Branch Report TR-08-01. (http://www.environmentyukon.gov.yk.ca/mapspublications/documents/peel_watershed_000.pdf)
Alberta	Biodiversity monitoring	Alberta Research Council. 2008. Integrated landscape management: Biodiversity monitoring. Alberta Biodiversity Monitoring Institute. (http://www.arc.ab.ca/areas-of-focus/integrated-landscape-management/biodiversity-monitoring/)
	Forest-based regional classification system	Alberta Sustainable Resource Development. 2003. Ecological land survey site description manual. 2nd ed. Resource Data Branch. Strategic Corporate Services Division. Pub No: T/036. ISBN: 2819-7.
	Field guide	Archibald, JH and Beckingham, JD. 1996. Field guide to ecosites of northern Alberta. Special Report 5. Canadian Forest Service. Northwest Region. Northern Forestry Centre.
	Natural resource management	Government of Alberta. 2009. Ecological land classification. Sustainable Resource Development, October 2009. (http://www.srd.alberta.ca/MapsFormsPublications/Maps/ResourceDataProductCatalogue/Ecological.aspx)

Northwest Territories	Biodiversity monitoring	Government of Northwest Territories. 2005. Ecosystem classification of the Northwest Territories. Environment and Natural Resources. (http://www.enr.gov.nt.ca/_live/pages/wpPages/Ecosystem_Classification.aspx)
	Field guide	Ecosystem Classification of the Northwest Territories (currently under development). The NWT Ecosystem Classification Project was initiated in 2004 with the objective of revising the ecozones and ecoregions defined under the Canadian National Ecosystem Framework. Contact: Mr. Bob Decker, Forest Management Division, Environment and Natural Resources. Government of the Northwest Territories. 867.874.2009.
Saskatchewan	General information	Saskatchewan Conservation Data Centre. 2009. Ecoregions of Saskatchewan. Saskatchewan Ministry of Environment. (http://www.biodiversity.sk.ca/eco.htm)
		Virtual Saskatchewan. 2007. Interactive Ecoregions Map. (http://www.virtualsk.com/maps/ecoregions.html)
Prairies	Field guide	AAFC (Agriculture and Agri-Food Canada). 1995. Terrestrial ecozones, ecoregions and ecodistricts, Alberta Saskatchewan and Manitoba, Canada. Canadian Soil Information System: State of the Environment Directorate, Ecozone Analysis Branch, Ottawa.
Manitoba	Biodiversity monitoring	Walker D, Barber D, Baydack R, Campbell, M. 2002. Manitoba ecosite classification and decision support system. University of Manitoba. (http://www.umanitoba.ca/geography/ecosite_pages/index.html)
	Forest-based regional classification system	Manitoba Conservation. 2001. Introducing Manitoba's forests. Government of Manitoba, Forestry Branch. (http://www.gov.mb.ca/conservation/forestry/forest-education/general.html)
Ontario	Biodiversity monitoring	Government on Ontario. 2005. Ontario biodiversity strategy – Working draft. Ministry of Natural Resources, March 2005. (http://www.obs-sbo.ca)
	Forest-based regional classification system	Government of Ontario. 2006. Forest resources of Ontario. Ministry of Natural Resources, Forests Division. June, 2006. (http://www.mnr.gov.on.ca/en/Business/Forests/Publication/MNR_E005106P.html)
	Field guide	Lee H, Bakowsky W, Riley J, Bowles J, Puddister M, Uhlig P, McMurry S. 1998. Ecological land classification for southern Ontario: First approximation and its application. SCSS Field Guide FG-02, September 1998. Federation of Ontario Naturalists.

		Government of Ontario. 2007. Ecological land classification primer. Ministry of Natural Resources, Policy and Planning Coordination Branch, March 30, 2007. (http://www.mnr.gov.on.ca/264777.pdf)
		Sims RA, Towill WD, Baldwin KA, Wickware, GM. 1997. Field guide to the forest ecosystem classification for northwestern Ontario, 2nd ed. Field Guide FG-03. Forestry Canada and Ontario Ministry of Natural Resources, Thunder Bay, Ontario, pp. 176.
	Regional land use planning/ conservation	Credit Valley Conservation Authority. 2010. Ecological land classification. Credit Valley Conservation Authority, Natural Heritage. (http://www.creditvalleycons.com/programsandservices/downloads/ELCwebsite.pdf)
		Rideau Valley Conservation Authority. 2010. Ecological land classification. Watershed Information Site, Aquatic Habitat and Terrestrial Ecology. (http://www.rvca.ca/watershed/aquatic_habitat/eco_overview.html)
Québec	Regional land use planning/ conservation	Gouvernement du Québec. 2003. Ecological land classification hierarchy. Ministère des Ressources naturelles, de la Faune et des Parcs 2003-3064. (http://mrnf.gouv.qc.ca/english/publications/forest/publications/ecological.pdf)
New Brunswick	General information	New Brunswick Department of Natural Resources, Ecological Classification Working Group. 2007. Our landscape heritage: The story of ecological land classification in New Brunswick. 2nd edition, ISBN 978-1-55396-203-8, pp. 359. (http://www.gnb.ca/0399/OurLandscapeHeritage/)
Nova Scotia	Regional land use planning/ conservation	Neily PD, Quigley E, Benjamin L, Stewart B, Duke T. 2003. Ecological land classification for Nova Scotia: Volume 1 – Mapping Nova Scotia's terrestrial ecosystems. Nova Scotia Department of Natural Resources, April 2003 Report DNR 2003-2. (http://www.gov.ns.ca/natr/forestry/ecological/ecolandclass.asp)
Prince Edward Island	Forest based regional classification system	Government of Prince Edward Island. 2009. Forest cover mapping. Department of Environment, Energy and Forestry. (http://www.gov.pe.ca/envengfor/index.php3?number=49737&lang=E)
Newfoundland and Labrador	Forest-based regional classification system	Government of Newfoundland and Labrador—Canada. 2010. Ecoregions of Newfoundland. Natural Resources, Forest Services Branch. (http://www.nr.gov.nl.ca/forestry/maps/eco_nf.stm)

Government of Newfoundland and Labrador—Canada. 2010. Ecoregions of Labrador. Newfoundland. Natural Resources, Forest Services Branch. (http://www.nr.gov.nl.ca/forestry/maps/eco_lab.stm)

Appendix C

Case Studies for Selecting Sampling Strategies and Calculating the Number of Samples

C.1 Introduction

This appendix contains several simplified scenarios that illustrate some of the issues involved with developing a sampling strategy and are meant to augment the description of sampling strategies presented in Subsection 3.3.5. The concepts illustrated in the scenarios are generally applicable to similar real-world sampling strategies. The case studies use simplified datasets to illustrate how to calculate the number of samples that need to be collected for each case study.

The variety of scenarios regarding soil sampling for the purpose of biological toxicity testing is presented, to the extent possible, in a hierarchical manner. That is, some of the simpler scenarios can and should be nested within a more complicated scenario. For example, consider the following possible scenario:

> A pipeline containing a deleterious substance ruptures. The pipeline ruptures at a point where it begins to traverse a valley. The substance release follows the topography downhill, pooling and spreading out laterally. There are two distinctly different receiving environments; the first is the slope, which for the purposes of this example is gravelled and steep. The second receiving environment at the base of the slope is comprised of much finer materials (e.g., silt and/clay) and a different flora due to the elevated water content of the downslope plain. Depending upon the nature of the substance released, there may be interest in:
>
> - delineating certain toxicity isopleths;
> - understanding the "average" or general toxicity within a prescribed area at the bottom of the slope; and,
> - understanding the toxicity gradient down the slope, which may be especially relevant if the substance released has physical properties such that the sloped receiving environment functions as a fractionating column.

Once the toxicity test responses are "understood" to the appropriate degree it may be of interest to:

- compare the results of toxicity tests for one, or both, distinct receiving environments with toxicity test responses in similar but uncontaminated soil;
- estimate a threshold effect for organism responses that represents a meaningful level of ecological functionality/dysfunctionality, or a level that triggers a management response; and,
- determine if the down-slope receiving environment is heterogeneous with respect to factors that either cause differential sequestration or degradation of the contaminant(s), thereby directly affecting toxicity by affecting contaminant bioavailability or indirectly affecting the test organism response by providing a more, or less, suitable micro-habitat.

Statistical sampling strategies should be developed in consideration of the receiving environment (there are a host of potential considerations falling into two general classes, factors that affect the toxicity and factors that indirectly affect the response) and intended data usage. A sampling strategy that may be appropriate on the basis of intended data usage may be inappropriate due to factors in the receiving environment and vice versa. As toxicity testing of soils (as well as chemical characterization) can be costly, it is in the best interest of all stakeholders to thoroughly discuss intended data usage early in the contaminated site assessment process, and then in consideration of a specific site, develop a cost-effective sampling strategy.

C.1.1 Notes to Readers

C.1.1.1 Correction for Finite Populations

In the unlikely event that a significant proportion of the soil on a site is sampled special statistical considerations are warranted. The following discussion, although technical in nature, should be understood by all using the case studies as guidance when sampling soils for the purpose of conducting biological tests.

When sampling a population comprised of discrete units such as trees within a woodlot there is a finite sample size, N. As the number of trees sampled, *(n)* increases, the ratio *n*/N will approach one. As this ratio approaches one, variance estimates are affected. In fact if *n*/N = 1 then the variance of the estimated mean should be zero because there is no uncertainty (ignoring measurement error) regarding the mean, which is now by virtue of sampling the complete population, the population mean. Correction for a finite population is often ignored as it has negligible effects on the estimates when the ratio is small.

When sampling contaminated sites using soil toxicity tests, the volume of soil sampled relative to the total amount of soil in the area of interest (the population) is likely to be very low. Corrections for a finite population are omitted from the following examples because:

- as discussed, the ratio of soil sampled relative to soil in the population is likely to be very low and therefore a correction will have negligible effects;
- to a lesser extent the idea that the concept of a discrete soil "entity" is a nebulous concept; and,
- there is a desire to simplify calculations.

If a substantial volume of soil relative to the total amount of soil at the site is collected for toxicity testing, a text such as that by Valiant *et al.*, (2000) or a statistician should be consulted.

C.1.1.2 How to Use the Guidance in this Appendix

This section presents a variety of case studies that augment the description of sampling strategies presented in Subsection 3.3.5. These case studies do not explicitly discuss why a specific sampling strategy was used, as guidance on selecting a sampling strategy is provided in Subsection 3.3.5. In describing the scenarios that illustrate some of the issues involved with developing a sampling strategy some material is provided that is necessary for completeness sake, but not critical to the average reader. This material is flagged by a "detour" sign. Readers should exercise caution in these areas! Practical tips for readers are emphasized using italics. The symbols, acronyms and conventions used within this section are presented in Table C1. Statistical tables necessary to complete the calculations presented in this appendix are provided in Appendix J.

Table C.1 List of symbols, acronyms and conventions

Symbol/ Acronym/ Convention	Definition or explanation
α	type I error rate
β	type II error rate
v	degrees of freedom
Ha	alternative hypothesis
Ho	null hypothesis
h	a specific stratum in a stratified sampling plan
J	the number of systematic samples within a stratum
L	the number of strata in a stratified sampling plan
n	sample size
n_h	stratum specific sample size
s	standard deviation
s^2	variance
$t_{\alpha,v}$	a quantile from the t-distribution with level of significance = 100 - α% on v degrees of freedom
$x_{h,i}$	the i^{th} observation within the h^{th} stratum
$\overline{X}$	the sample mean
$\overline{X}_h$	the stratum specific sample mean

C.2 Case Study #1: Simple Random Sampling and the Control-Exposure Study

This is the simplest case study described herein and it often forms the basis of more complicated sampling strategies and sample size calculations. This study uses probability-based random sampling to protect against biases in sample location selection. The advantages and disadvantages of simple random sampling are discussed in Subsection 3.3.5.

C.2.1 Control Exposure Studies

Soil toxicity tests may be used to assess potential effects at an exposure site relative to a control site. The results may be used in environmental management decision-making. These types of studies are referred to as control-impact or control-exposure studies in the ecotoxicological literature.

A control-exposure study compares the mean of a set of replicate toxicity test results from an exposure "site" versus another control site. Both sites must be as similar as possible with respect to all factors that can affect the outcome of the toxicity test, with the exception of the contaminant(s) of concern. While this is a simple concept, as discussed in Subsection 3.6.2, there can be a great deal of debate regarding the conclusions following a comparison of effects at an exposure site to those at a control site. It is well worth the effort to ensure that a defensible control site is selected. Note that in the case of contaminants that are naturally occurring such as metals, it is not necessary to find a site that is absent of metals, only one that represents the local natural background. Details of estimating local background concentrations are found in Breckenridge and Crockett (1995), USEPA (2007), and to a lesser extent, Yukon Environment (2002).

It is critical that the exposure site of interest be defined *prior* to examining the data that will be used to test hypotheses. One common error when conducting soil chemical surveys is to collect the data, examine it, segregate a portion of the data as "hotspot" and then test a hypothesis that a chemical concentration is elevated in the hotspot. Another example occurs when an *a priori* decision is made to randomly collect a

number of replicates within an exposure site and use the mean to make a management decision if some threshold is not exceeded. If the mean is less than the management threshold but some of the replicates exceed the threshold it is incorrect to separate a subset of the area using these data and label this sub-area as exceeding the threshold. These two examples of incorrect data usage are often referred to as "data snooping." The practice of data snooping or using data to generate hypotheses and then using the same data to test hypotheses leads to incorrect confidence limits and a possibly incorrect conclusion about the site of interest. Note that the process of using initial or screening data to drive additional data acquisition is not data snooping as no hypothesis is being tested using a selection of observations.

Once an exposure site has been defined by whatever means and a suitable control site is chosen, the procedure for selecting sampling locations within the exposure site and the control site must be selected.

In this case study, the assumption is made that the exposure site is homogenous with respect to variables that can affect the toxicity test result(s). This assumption is relaxed in subsequent case-studies. Given the assumption of homogeneity of factors that may affect the toxicity test results, a simple random sampling scheme is appropriate within the designated sites.

C.2.2 Implementing the Random Sampling Strategy

Soil for the purposes of this example is a continuum in 2-dimensions. The easiest method to select the *"n"* sampling locations is, by drawing a rectangle on a map that represents the x and y coordinates within which the site is contained. Using a uniform random number generator or tables of random numbers[36], choose x and y locations at a level of resolution that is consistent with the positioning methods used for the study. If some of the selected locations fall outside an irregularly spaced area continue choosing random samples until *n* locations within the sampling site are selected.

C.2.3 One and Two-Sided Hypothesis Tests

Before estimating sample sizes, the nature of the difference in toxicity test results between control and exposure sites must be considered. In some cases *any difference* between results at the control and exposure sites is undesirable. Thus, if the mean exposure site response is higher *or* lower than the control response a deleterious effect has occurred. When testing hypotheses this is referred to as a two-tailed test. In other circumstances *only* a decrease in the mean exposure response relative to mean control response is considered deleterious. Thus, an increase in the mean exposure response relative to the mean control response is of no concern. When testing hypotheses, this is referred to as a one-tailed test.

Given a fixed sample size a one-tailed test is more powerful than a two-tailed test and should always be used when it makes sense to do so. However, the decision in selecting between a one and two-tailed test must be made prior to examining data. If the decision is made after examining the data, this is a form of data snooping and it will lead to incorrect error rates.

C.2.4 Choosing a Sample Size

A sample size is a function of the desired type I and II error rates (see Subsection 3.3.10) and is driven by the variability of the toxicity test results and the minimum difference in mean toxicity test results between the control and exposure sites that are of interest. The greater the variability in toxicity test results and the smaller difference that it is important to detect, the larger a sample size is required to achieve desired type I and II error rates. At times the variability may be so large that the sample sizes required to achieve the desire type I and II error rates are impractically large. If this is the case, it might be worth also considering whether some unsuspected variable is confounding the results. If yes, an additional level of stratification may be required to subdivide the site into homogenous sub-sites, or, statistical tools might be used to control for the effects of confounding variables, possibly in consultation with a statistician. If no, a larger

---

[36] Available in statistical textbooks such as Sokal and Rohlf (1995) and Zar (1999).

detectable difference between exposure and control means might be selected or the type I and II error rates might be increased. However, see Subsection 3.3.10 for a discussion regarding type II error rates when environmental protection is predicated upon not rejecting the null hypothesis.

Prior to conducting sample size calculations some estimate of the variability of the toxicity test response must be obtained. This estimate might use data from similar experiments or best professional judgment; consultation with personnel from an ecotoxicity testing laboratory for this information is advisable. The estimate of variability represents the "population" variance, which may be estimated by pooling variances that represent the same species and exposure conditions.

Example 1 is used to illustrate the decisions and calculations required to estimate sample sizes for a simple random sampling scheme used to test hypotheses of the control-exposure type.

C.2.4.1 Example 1 Considerations.

This example continues with the scenario presented at the beginning of this appendix; that of a leak in a pipeline where the spill runs down a slope and pools at the bottom. In this example interest lies in assessing soil toxicity at the base of the hill to see whether there is a deleterious effect relative to a nearby, suitable reference soil. A test in which the response is seedling root growth has been chosen for reasons of taxonomic representativeness and sensitivity to the weathered hydrocarbons present. Specific considerations required to estimate sample sizes are:

- Difference between means that is important to detect = 20 mm root length.
- Only a decrease in root length is considered deleterious; therefore, a one-sided test is appropriate.
- Two sample standard deviations are available[37]:
  - One is from a nearby study with a standard deviation, $s_1 = 12$ with a sample size, $n_1 = 14$; and,
  - Another available standard deviation, $s_2 = 10$ with a sample size, $n_2 = 10$.
- The type I and II error rates are both[38] 10%.

C.2.4.2 Example 1 Calculations

Using the available estimates of standard deviations, the pooled variance s_p^2 may be estimated using:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)} \qquad [1]$$

where:

- s_p^2 is the pooled variance; and,
- s_i^2 and n_i are the variance and sample size corresponding to dataset "i".

Using the Example 1 data above,

---

[37] While it is most desirable to obtain variances from pilot studies or in the first step of a multi-step sampling program, these variances may not be available. Variances for planning purposes may be obtained from studies that are as similar as possible, from the literature or using expert judgement. Note that the idea of sampling sufficiently to meet data quality objectives is discussed in Subsections 3.3.1 and 3.3.5.

[38] In Canada, type I and II error rates are often set equal to one another when making environmental decisions. This places equal risk on unnecessary action and missed action.

$$s_p^2 = \frac{(14-1)144+(10-1)100}{(14-1)+(10-1)}$$

$$= \frac{2772}{22} \qquad [2]$$

$$= 126$$

The required sample size for a two-sided test[39] is estimated as:

$$n \geq \frac{2s_p^2}{\delta^2}\left(t_{\alpha/2,\nu} + t_{\beta,\nu}\right)^2 \qquad [3]$$

where:

- n is the number of samples collected from each of the control and exposure areas;
- s_p^2 is the pooled variance;
- δ is the difference of interest;
- $t_{\alpha/2,\nu}$ is a quantile from the t-distribution with α = type I error rate on ν degrees of freedom; and,
- $t_{\beta,\nu}$ is a quantile from the t-distribution with β = type II error rate on ν degrees of freedom; and,
- ν, the degrees of freedom are equal to $2(n - 1)$.

For a one-sided test change $t_{\alpha/2,\nu}$ to $t_{\alpha,\nu}$

Using the Example 1 data and considerations above, we are faced with an immediate challenge. How can we obtain the quantiles from the t-distribution (from the t-tables in Table J.2 and J.3; Appendix J) required in the sample size calculation formulae without knowing the sample size? We guess at the required sample size and estimate the sample size using the formula. In step 2 we use the estimated sample size as degrees of freedom for the t-distribution and re-estimate the required sample size. This iterative procedure continues until there is no change in the estimated sample size for two successive iterations.

There are various software routines for estimating sample sizes; if estimating sample sizes manually, it is better to start with a larger than a smaller initial guess. In this case we will guess that 7 samples are required from each of the control and exposure sites. Thus ν, the degrees of freedom are equal to 2 (7 - 1) = 12 and recalling that a one-sided test is being used, $t_{0.10,12} = 1.356$.

$$n \geq \frac{2\times126}{20^2}(1.356+1.356)^2 \qquad [4]$$

$$\geq 4.6336$$

[39] Modification for a one-sided test follows immediately below definitions.

Erring on the side of conservatism we chose $n = 5$ as the number of samples required for each of the control and exposure sites. Repeating the previous calculations using the "guess" $n = 7$, with $v = 2(5\text{-}1) = 8$; $t_{0.10,8} = 1.397$ we obtain $n \geq 4.9181$. Again, erring on the side of conservatism we chose $n = 5$ as the number of samples required for each of the control and exposure sites. Since the sample size estimates have produced the same estimated samples on two successive iterations we chose a final sample size of $n = 5$ samples from each of the control and exposure sites.

C.2.4.3 Validity of Sample Size Calculations.

For the sample sizes estimated using these methods it is assumed that the toxicity test endpoints follow a Normal or Gaussian distribution and that each toxicity test endpoint is independent of another. Because toxicity test endpoints are themselves estimated from toxicity test responses, the process of estimating the endpoint may induce[40] a normal distribution among the endpoints.

If there is doubt as to whether the distribution of toxicity test endpoints is normal then a non-parametric test will likely be used to test the hypothesis of equality of medians between control and exposure sites. As the non-parametric equivalent of the *t*-test, the Mann-Whitney test is less powerful; therefore, the sample size necessary to achieve the same quality of test must be adjusted upwards by multiplying the estimated sample size by 1.2.[41]

Another common assumption of frequently employed statistical tools (tests of hypotheses and sample size calculations) is that observations are independent of one another. At this point in time it is not clear whether this assumption will be met for soil toxicity test responses. If it is not met and the correlation structure (which will almost certainly be positive) is substantive then sample sizes will be underestimated and hypothesis test levels of significance will be incorrect.

The independence of replicate toxicity test results within an area may be difficult to assess given the expected relatively small sample sizes. *If there is correlation evident among ancillary data such as chemical concentrations, a statistician should be consulted.*

C.2.4.4 Making Inferences.

Once the data are collected and quality assured, the data from the control and exposure sites may be compared. Following Example 1 and using the raw data provided in Table C.2, the null hypothesis, Ho: The average toxicity test response in the exposure area is the same as that of the control area is tested against the alternative hypothesis, Ha: The average toxicity test response in the exposure area is less than that of the control area. Details of conducting *t*-tests including the requisite tests of assumptions may be found in introductory statistics textbooks. Two such textbooks used by biologists are Sokal and Rohlf (1995) and Zar (1999). Results of a *t*-test assuming equal variances are presented below.

The *t*-test test statistic for the example dataset is 2.323 on 8 degrees of freedom with a *P*-value of 0.0243. The very small *P*-value suggests that there is sufficient evidence to reject the null hypothesis and accept the alternative. We conclude that the mean root length in the exposure area is less than that of the control area at a $(1\text{–}0.0243) \times 100\% = 97.57\%$ level of significance.

In the event that the data did not pass the test of normality, the nonparametric Mann-Whitney test may have been used to compare the exposure and reference medians.

---

[40] Theorem 18: Mood *et al.* 1974.

[41] Based on the asymptotic relative efficiency of the Mann-Whitney test relative to the two sample Student's *t*-test as discussed in Nöether (1987).

C.2.5 Additional Considerations

The following refinements of basic sampling plans for control-exposure type comparisons are beyond the scope of this document but are included to bring the possibilities to the attention of field practitioners. Advice from a statistician may be necessary to implement these refinements.

C.2.5.1 Heterogeneous Variances.

In Example 1, an equal allocation of sampling effort in the control and exposure sites was used. This practice may be contraindicated when variability among the two sites is dissimilar. In this case, disproportionately allocating sampling effort can reduce the total number of samples, for fixed type I and II errors and difference of interest.

C.2.5.2 Correlated Data.

Observations are autocorrelated or correlated in space with one another if observations closer together in space are more similar than those further apart. When data are autocorrelated and the correlation is positive (which it will almost certainly be in the case of soil toxicity tests if autocorrelation exists) variance estimates will be biased downwards. This will artificially decrease sample sizes and incorrectly increase the power of statistical tests. To some extent random sampling guards against these effects; however, a stratified random sampling (StRS) plan will generally estimate a smaller variance for the overall mean than a simple random sampling (SRS) plan. If autocorrelation is suspected, a statistician should be consulted.

C.2.5.3 Cost Considerations.

At times it may be more costly to sample one area relative to another due to site accessibility, etc. In this case, disproportionate sampling among areas may be considered to optimize cost.

Table C.2 Example 1 data set

	Root lengths (mm)	
	Control	**Exposure**
	37.6	11.3
	29.7	7.0
	36.8	4.9
	51.1	17.4
	55.1	9.1
Estimated mean	42.06	9.93
Estimated standard deviation	10.63	4.82

C.3 Case Study #2 – A Stratified Random Sample Study

Stratified random sampling is appropriate when known factors that affect the toxicity test response are present and there is interest on estimating a parameter over all strata. For example, if the site investigator is interested in estimating parameters within a stratum then an alternative sampling program within a stratum is appropriate. The failure to control for a known source of variability can lead to needless sampling. The formulae in this section are adapted from Cochran, (1977). Adaptations have been made to simplify the generally applicable formulae for the benefit of the reader.

One adaptation deals with non-discrete nature of soil where an individual sampling unit and hence "sampling frame" cannot be defined. Thus stratum weights defined as the ratio of within-stratum sampling units relative to the total number of sampling units in the population can only be defined as 1) proportions

based on surface area, or 2) a conceptual sampling unit based on the amount of soil that a sample represents. This latter approach is conceptually unsatisfactory because the overlap of sampling units due to the physical nature of collecting a soil sample contravenes the concept that sampling units are mutually exclusive (a sampling unit is unique and may only be sampled once).

This stratified random sampling example continues with the scenario of a pipeline break with the loss of a deleterious substance down a hill slope. After completing the first study using root lengths in an area at the base of the hill where hydrocarbons had pooled it was noticed that toxicity was greater (as manifested by decreased root length measurements) along a band located, not directly at the base of the hill, but a bit further away from the hill and parallel to the slope face. Investigation of soil physical properties showed that soil in this area of the site had a higher proportion of coarse particle sizes. In accordance with the study objective (not described in this example), the entire area at the base of the hill is being assessed for toxicity rather than a portion thereof. Thus an assessment of the overall site mean toxicity test response is required.

C.3.1 Implementing the Stratified Sampling Scheme

Once strata have been chosen, samples from within each stratum must be collected. Since by definition, the area within a stratum should be as homogeneous as possible, random sampling within strata is again appropriate. Methods for physically selecting random samples are the same as those presented in subsection C.2.2.

C.3.2 Estimating the Overall Mean and Variance

A discussion of how many samples should be chosen overall or how to optimally allocate samples within strata, although logically placed here, is deferred in order to introduce some of the necessary statistical nomenclature and concepts in a stepwise fashion.

The first concept is estimating the overall mean, for example, shoot length. This overall mean is estimated using a weighted mean of shoot lengths from each of the two strata. The stratum-specific mean is estimated using the usual formula for a mean as follows:

$$\bar{x}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi} \qquad [5]$$

where:

- h refers to a specific stratum;
- $\bar{x}_h$ is the stratum specific mean;
- n_h is the stratum specific sample size; and,
- $x_{h,i}$ is the i^{th} observation within the h^{th} stratum.

Now, the overall mean across strata $\bar{x}_{str}$ is estimated using:

$$\bar{x}_{str} = \sum_{h=1}^{L} W_h \bar{x}_h \qquad [6]$$

where:

- L is the number of strata; and,
- W_h is the stratum weight. For the *purposes of this document*, stratum weights refer to the proportion of the total area for which a statistic is being estimated that a specific stratum h, represents.

The variance of the overall mean, $s^2(\bar{x}_{str})$ assuming the site (the population) being studied is very large relative to the number of samples, collected is given by:

$$s^2(\bar{x}_{str}) = \sum_{h=1}^{L} \frac{W_h^2 s_h^2}{n_h} \quad [7]$$

and where:

- s_h^2 the within-stratum variance is defined by:

$$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} \left(x_{h,i} - \bar{x}_h\right)^2 \quad [8]$$

Equation 7 demonstrates the benefit of stratified random sampling relative to simple random sampling, which is that the variance of the overall site mean is a function of only the within-stratum variances. Thus if strata are carefully selected, stratified random sampling can provide a markedly better (smaller variance) estimate of the mean than SRS.

C.3.3 Optimally Allocating Sampling Effort within Strata

The simplest formula for allocating sampling effort assumes that the cost of collecting samples from each stratum is the same and that interest centers on minimizing variability for a fixed cost (a pre-determined total sample size, n). In this case the optimal (Neyman) allocation (Neyman, 1934) is:

$$n_h = n \frac{W_h s_h}{\sum_{h=1}^{L} W_h s_h} \quad [9]$$

Costs can vary by strata if for example, sub-site access is an issue. In this case, Cochran (1977) presents the following formula for the optimal allocation of sampling effort when sample size is fixed and interest is on minimizing variability:

$$n_h = n \frac{W_h s_h / \sqrt{c_h}}{\sum_{h=1}^{L} W_h s_h / \sqrt{c_h}}, \quad [10]$$

where:

- c_h = cost of sampling in the hth stratum.

This allocation assumes that the cost of collecting samples is given by the following:

$$\text{cost} = c_0 + \sum_{h=1}^{L} c_h n_h, \quad [11]$$

where:

- c_0 is the overhead cost. This might be the cost of travelling to the site, etc.

C.3.3.1 Example 2 Considerations.
Within the previously defined site of interest, the band of coarser materials comprises 40% of the total site. The same type I and II errors apply as do other specific considerations except that now shoot lengths will be used to assess soil toxicity. The difference that it is important to detect is 20 mm. The cost of collecting a sample is the same within each stratum and, given the project budget there are sufficient financial resources to collect and analyze $n = 24$ samples. The question solved in the next question is "What is the best way to allocate these 24 samples"?

C.3.3.2 Example 2 Calculations.
This section presents calculations demonstrating how to optimally allocate sampling effort for a fixed cost. Equation 9 is used for this purpose. W_h, the stratum-specific proportions of the total area are 0.40 and 0.60 for the L = 2 strata. A total of $n = 24$ samples will be collected.

Table C.3 Preliminary calculations demonstrating allocation of effort in StRS

Stratum_h	W_h	s_h	$W_h s_h$
1	0.40	8	3.20
2	0.60	4	2.40
			5.60

$= \sum_{h=1}^{L} W_h s_h$

Using Equation 9 for stratum 1,

$$n_1 = 24\frac{0.4\times 8}{5.60} = 13.71 \qquad [12]$$

Similarly, for stratum 2 $n_2 = 10.29$. Thus the optimal allocation of samples is 14 to stratum 1 and 10 to stratum 2. This example illustrates an important aspect of StRS, that the optimal allocation of samples is strongly influenced by the degree of within-stratum variability. *All other considerations (cost, W_h, etc.) being equal, it is always advantageous to collect a proportionally higher number of samples from strata exhibiting higher variability.*

C.3.4 Estimating the Sample Size with a Pre-specified Variance
In the ideal situation, the total sample size will be driven by the need to acquire sufficient data to make a conclusion with a pre-specified, stakeholder acceptable level of precision rather than being potentially limited by cost. The formula for estimating sample sizes with a pre-specified variance is:

$$n = \frac{\left(\sum_{h=1}^{L} W_h s_h \sqrt{c_h}\right)\sum_{h=1}^{L}\left(W_h s_h / \sqrt{c_h}\right)}{V + \frac{1}{N}\sum_{h=1}^{L} W_h s_h^2} \qquad [13]$$

where:

- c_h is the cost of sampling the h^{th} stratum;
- V is the desired variance; and,
- N is the number of sampling units in the population.

C.3.4.1 Example 3 Considerations.
The data of Example 2 are used to illustrate sample size calculations using an example pre-specified variance = 12 mm^2. Additionally, in this example, each sample collected comprises a surficial area[42] of 0.8 m^2 and the total area of the site is 1200 m^2.

C.3.4.2 Example 3 Calculations.
Equation 13 is used to estimate the total sample size as follows:

- L, the number of strata is 2;
- N, the number of sampling units in the population is = total area/surficial area per sampling unit = 1200 m^2/0.8 m^2 = 1500;
- W_h, the stratum weight and s_h, the within- stratum standard deviations are presented in Table C.4;
- c_h is constant among strata and is = $225; and,
- V, the desired variance is 12 mm^2.

Table C.4 Preliminary calculations demonstrating sample size calculations for StRS

Stratum$_h$	W_h	s_h (mm)	c_h ($)	$W_h s_h \sqrt{c_h}$	$W_h s_h / \sqrt{c_h}$	$W_h s_h^2$
1	0.40	8	225	72	0.32	57.6
2	0.60	4	225	126	0.56	117.6
			Sums over strata	**198**	**0.88**	**175.2**

Now,

$$n = \frac{198 \times 0.88}{12 + \frac{1}{1500} 175.2} \qquad [14]$$
$$= 14.38$$

Erring on the side of caution, this number is rounded up to 15. These fifteen samples may be allocated among the strata proportionately, or using the optimal Neyman allocation presented in Equation 10 (Equation 9 could also be used to optimally allocate the fifteen samples).

[42] This example illustrates one of the challenges encountered when estimating sample sizes for application in a medium such as soil that is continuous. The sample size formulae, presented in this document use classic Horvitz-Thompson (1952) estimators, which assume a finite population consisting of distinct sampling units. This assumption is not met in a continuum such as soil because a distinct sampling entity does not exist. The solution provided creates a hypothetical distinct entity, the "sample" collected. In the example, the surficial area that a sample (sampling unit) represents can vary due to differences in the depth of the horizon(s) of interest over the site. Thus, the estimated sample sizes will only be approximate. Cordy (1993) discusses extensions of the Horvitz Thompson estimators to sampling from a continuum. This is an area of active research; field personnel concerned with this issue are encouraged to consult a statistician.

C.3.5 Correlated Data

The variance of the overall mean when data are correlated is better estimated using a StRS plan than a SRS plan.

C.3.6 Suggestions for Stratification

One of the most obvious criteria for stratification in systematic sampling programs is the concentration of the contaminant(s) of concern. Pennock *et al.* (2008, Table 1.2) classifies the variability of various soil properties as low, moderate, high, and very high. As some of these soil properties can affect contaminant concentration, modify the bioavailability of contaminants and also directly affect toxicity test responses, the variability of the global mean estimated using a stratified random sampling plan is a function of the within-stratum variability soil properties (see Equation 7) should be considered as stratification variables.

C.4 Case Study #3 – A Transect Study

A systematic linear sampling (SLS) plan collects a number, k, samples at pre-specified increments along a transect. The initial starting point is chosen randomly. If all samples are collected on the basis of this single randomly selected starting point, the systematic linear sampling plan consists of one sample from the population, with k subsamples. This is commonly known as a "1-in-k" systematic sample. A "1-in-k" systematic sample provides an unbiased estimate of the population mean; that is, the estimated mean does not deviate systematically from the true, but unknown, population mean. The same is not true; however, of the estimated variance of the sample mean (Equation 1) because *only one random sample* has been drawn from the population (the site). This may seem counterintuitive to some readers since "k" physical samples have been collected. However, once the initial sampling location is randomly selected, all other sample locations are fixed in a systematic sampling scheme. This causes problems with respect to estimating the variance of the mean. Since this variance term is used to make inferences regarding the sample mean, and also to estimate sample sizes, it is important to obtain an unbiased estimate of the mean. Methods for obtaining an unbiased estimate of the mean for one sample collected along a transect are provided within this subsection (C.4) (methods for calculating the number of samples are of course moot as only one sample is collected per transect). Methods for implementing the basic transect study are presented in Subsection C.4.1.

Systematic sampling plans that collect only a single sample are problematic from a statistical perspective; however, they are often used by field practitioners due to ease of use and ensured coverage of a site. A common practice with this type of sampling strategy is, instead of randomly selecting the initial starting point, to select the first sample location close to a point source of pollution; *it is worthwhile to explicitly note that this is an incorrect practice as the non-random individual sample collected at a site can no longer provide an unbiased estimate of the population mean.*

One further problem that challenges systematic sampling plans is periodicity. Periodicities are often encountered when collecting data associated with some diurnal rhythm. It is unlikely that periodicities use will occur in most soil toxicity sampling applications. One possible exception may be in "land farms" used to bioremediate contaminated soils.

Finally, systematic studies may also be conducted in two-dimensions using a grid and a randomly chosen initial sample location.

C.4.1 Implementing the Transect Study

A single sample transect study is implemented by choosing the desired distance between sampling locations such that the desired number of sub-samples (discrete samples along the gradient) are collected over the length of the transect. Then, a random starting position along the transect is selected. Sub-samples are collected in one direction down the transect until the end of the transect is reached. Then samples are

collected from the beginning of the transect until the initial, randomly selected sampling location has been reached.

C.4.2 Estimating the Mean and Associated Variance of a Single Sample

The mean of a set of observations collected through a systematic sampling plan is estimated as one normally calculates a mean. However, an unbiased estimate of the variance of the mean (the standard error of the mean) is not straightforward to estimate for a transect study. The standard variance estimate is incorrect. Instead the following *approximate* method may be used if the population is in random order (there are no trends in the toxicity test response). Note that when trends in toxicity test responses occur (as will often be the case within a site being studied) the following formula should *not* be used. Instead the modified systematic designs presented in subsection C.4.2.1 (Alternative Systematic Designs) should be used. Geostatistical tools (not described herein) may also be used[43], likely in consultation with a statistician.

The *approximate* estimate of the variance of a mean using a single sample from a transect is:

$$\mathrm{var}(\bar{x}) = \frac{1}{n(n-1)} \cdot \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 \qquad [15]$$

where:

- n is the sample size;
- x_i is the i^{th} observation; and,
- $\bar{x}$ is the sample mean.

C.4.2.1 Additional Considerations: Alternative Systematic Designs

The following alternative systematic designs may be used to obtain unbiased variance estimates of the mean when using systematic sampling techniques.

Multiple Systematic Sampling. This method follows the single sample, systematic sampling plan presented as the original example in this Case Study but uses J multiple random starting locations. The formulae presented below are discussed in Gilbert (1987).

If the mean of the j^{th} systematic sample is the usual arithmetic mean $\bar{x}_j$, then the transect mean is

$$\bar{x} = \frac{1}{J} \cdot \sum_{j=1}^{J} \bar{x}_j \qquad [16]$$

and the variance of the mean[44] is

$$\mathrm{var}(\bar{x}) = \frac{1}{J(J-1)} \cdot \sum_{j=1}^{J} \left(\bar{x}_j - \bar{x}\right)^2 \qquad [17]$$

---

[43] McArthur (1987) simulated a two-dimensional Gaussian dispersed pollutant and concluded that the best sampling plans to estimate the mean of a locally concentrated pollutant are stratified systematic and then unstratified systematic with the sampling plan appropriate estimators. These estimators performed better than kriged estimates.

[44] A finite population correction factor is available for these formulae.

Systematic Stratified Sampling. In systematic stratified sampling (SSt), multiple (at least two) systematic samples are collected within each stratum. The mean and variance of stratum "h" can be estimated using Equations 15 and 16, respectively. Once the stratum-specific mean and variance are estimated the population mean and variance may be estimated using Equations 6 and 7, respectively.

C.4.2.2 Additional Considerations: Correlated Data.
Like the SRS, a Sy sampling plan will produce estimates of the variance of the mean that are biased (Flores *et al.*, 2003) when data are correlated. Whether this bias is less than or greater than that estimated using a StRS cannot be generalized on the basis of the strength of the autocorrelations.[45] A Sy sampling plan can produce results that are better or worse than a StRS plan. If the data are correlated, a statistician should be consulted.

C.5 Case Study #4 – Detecting an Area of Elevated Toxicity

Environmental management decisions using toxicity test results may involve detection and or delineation of areas in a site of elevated toxicity. Note that a "hot spot" so defined may differ than that defined by chemical concentrations alone as the biological toxicity test results integrate exposure to individual contaminants, additive, synergistic and/or antagonistic effects of individual contaminants, factors that affect bioavailability and factors that influence the biological test organism responses.

Sampling strategies useful for detecting hot spots include systematic sampling strategies, adaptive variants of other types of sampling strategies and in some instances composite sampling under other types of sampling strategies. The optimization of suitable sampling strategies varies with the intended data analytical approach (i.e. design-based which uses classic inferential approaches, and model-based which simultaneously models the spatial nature and of the response(s) and the associated correlation structure), intended data usage, the sampling approach (fixed in advance or adaptive), whether compositing is allowed, etc. This case study presents a common scenario: how many samples must be collected to ensure that a hot spot of a given size and orientation is detected? Other possible questions are: What is the probability that a hot spot of a given size has been missed? and, what is the probability of detecting a hot spot of a given size given a specific grid sampling design? The case study largely follows Gilbert (1987), which in turn follows Singer (1972).

C.5.1 Example 4 Considerations

The calculations assume that:

- The unknown hot spots are randomly distributed;
- A sample can be unmistakably designated as belonging to a hot spot or not;
- When viewed from above, the outline of the hot spot is elliptical or spherical; and,
- The area represented by the samples collected is small relative to the total area of the site.

Before estimating the required grid spacing to detect the hot spot, an acceptable probability (β) is specified for not finding a hot spot of the desired size. In keeping with hypothesis testing precedents, this probability should be less than 10% and certainly not more than 20%.

C.5.2 Example 4 Procedure

Calculate the shape S, of the hot spot as:

---

[45] Correlations are positive, data are anisotropic and untrended.

$$S = \frac{I}{L} \qquad [18]$$

where:

- I = length of short axis/2; and,
- L = length of the long axis.

Note that the shape of S ranges from 0 to 1. The necessary grid spacing is determined by reading off charts in Figures 3, 4, or 5 in Zirschky and Gilbert (1984) (for square and rectangular or triangular grids, respectively) using the likely value of S and the desired value for β.

If the data are to be interpreted using kriging methods, Yfantis *et al.* (1987) concluded that [based on the same criteria used by Olea, (1984)] the equilateral triangular design was slightly more efficient than a rectangular grid but that a hexagonal grid was more efficient than either when micro-scale variability is large relative to overall variability and the distance between sample points approaches the distance where observations are functionally uncorrelated.

As figures for hexagonal grids are unavailable for estimating grid spacing to meet the specifications of the sampling strategy (e.g., shape of the ellipse, β, etc.), a triangular design is chosen for the purpose of illustration.

As the shape of the hot spot is unknown, a relatively small value of S = I/L = 12 m/40 m = 0.3 is chosen following Zirschky and Gilbert (1984). Using, β = 0.10 and Figure 5 of Zirschky and Gilbert (1984), we find that no grid can satisfy these criteria. If instead, a less elongated ellipse is chosen with S = 12 m / 20 m = 0.6, Figure 5 of Zirschky and Gilbert (1984) may be used as follows:

- draw a horizontal line from β = 0.10; and,
- where this horizontal line intersects the S = 0.6 curve;
- draw a vertical line down to the abscissa ("x"-axis) and read of the value of "L/G" = 0.7
- now using the definition of G presented in Figure C.1, we find that since L/G = 0.7 and L (from our example) is 20, then G = 28.57 m.

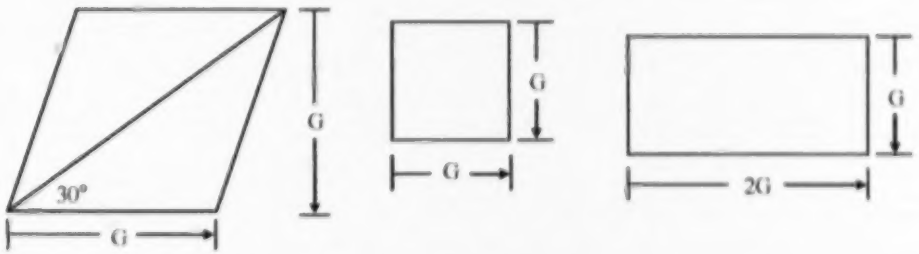


Figure C.1 Dimensions for (from left to right) triangular, square, and rectangular grids.

Geostatistics

D.1 Selection and Application of Variograms for Biological Assessment of Contaminated Sites

Biological assessments of soils are conducted to understand the biological responses to contaminated soils. As stated in Subsection 3.3.6, in order to generate spatial maps of the observed biological responses that acknowledge variability, geostatistical tools are used. These tools make certain assumptions about the process being modelled. The variogram may be used to assess these assumptions and better understand the geospatial biological response.

As described in Subsection 3.3.6, rather than use correlations, for historic reasons, geostatisticians use the covariance among observations to construct similar plots called variograms. The variogram describes the covariance among observations separated by a distance *"h"*. The variogram is estimated by

$$\hat{\gamma}(h) = \frac{1}{2N_h}\sum_{i=1}^{N_h}\left(Z(s_i) - Z(s_i + h)\right)^2 \qquad [19]$$

where:

- $\hat{\gamma}(h)$ is one half the average squared distance between observations separated by a distance *"h"*;
- $Z(s)$ is a random function (in this case a random toxicity test result);
- $Z(s_i)$ is the value of a function at location "i";
- h is a given distance; and,
- N_h is the number of sample pairs.

A generic variogram is presented in Figure D.1.

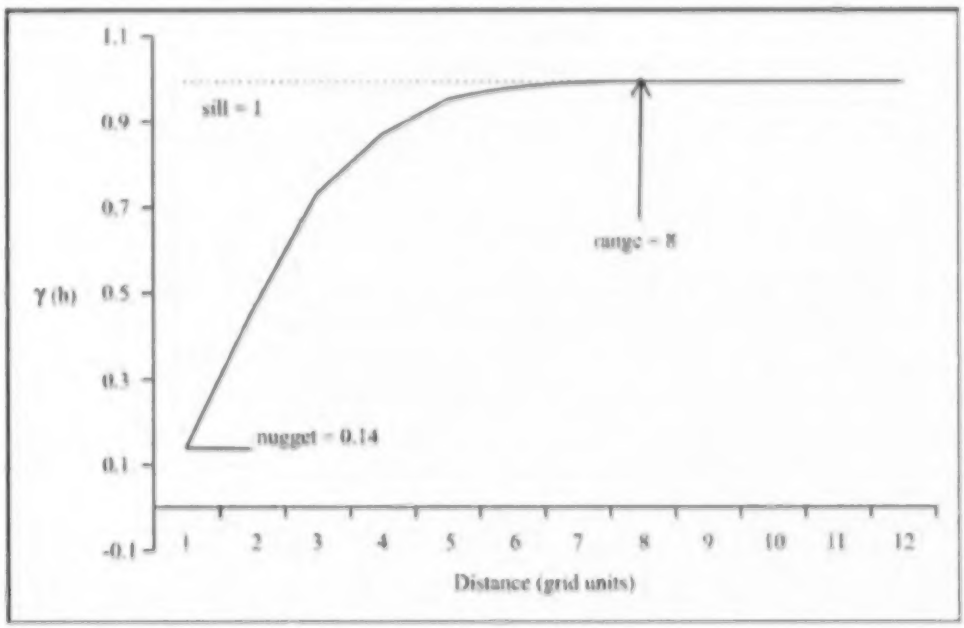


Figure D.1 Generic Variogram

In Figure D.1 the variogram has a horizontal asymptote[46] = 1 and is known as the sill. The point on x-axis at which the variogram "nears" the asymptote is called the range. Observations separated by more than the range are uncorrelated.

Although the correlation between observations separated by the theoretical distance of zero should be null, there is a small variation attributable to measurement or micro-scale variation known as the nugget. A large nugget effect relative to the sill may be of concern. However, there is an insufficient body of soil toxicity data to estimate what typical nugget effects may be and therefore what comprises a "large" nugget.

It is important to note that a variogram only describes the correlation structure among observations at a site and does not predict values. Prediction of values is conducted using the information provided by the variogram (modeled correlation among adjacent observations) using a linear interpolation technique called kriging (refer to Subsection 3.3.6.2 and D.2; Appendix D).

Some of the terminology associated with choosing and estimating variograms (presented below) is discussed motivated by a small idealized grid sampling program. Consider sample locations set out along a square grid where distance between transects, is *"h"*. As shown below, for any (non-boundary) location *"x"* there will be four other sampling locations that are exactly *h* units away.

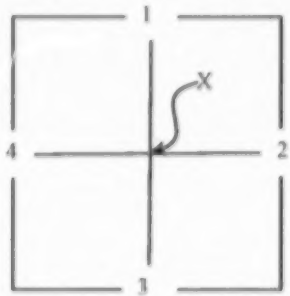


Figure D.2 Sampling Locations Separated by Distance *"h"* from Location *"X"*

Now consider a specific square grid consisting of three "east-west" and three "north-south" transects with a distance between transects = 1 metre (m). In this sampling scheme there will be 12 pairs of observations separated by 1 m and 6 pairs of observations separated by 2 metres. If we consider distances in the north-east, south-west and north-west, south-east directions other distances are possible. The various distances and number of possible pairs are shown in Table D.1.

In this example, due to the grid sampling, we can estimate the variogram at *only* 5 distances (1, 1.41, 2, 2.24, and 2.83 m).

If instead of speaking of specific distances we partition the distances into bins or ranges as is done when creating frequency histograms, we might create bins at 0.5-m increments. Thus "bin 1" or "lag 1"would correspond to observations separated by a physical distance of 0 to 0.5 m. In the grid sampling experiment presented, the number of pairs separated by this distance is zero. Bin 2 or lag 2 contains pairs of observations separated by 0.5 to 1.0 m and contains 12 pairs of observations. Subsequent bins or lags are defined similarly. The covariance is estimated for the pairs of observations falling into each bin and plotted against distance or lag number.

---

[46] A line that is approached but never reached. May be thought of as a limit. For the more mathematically inclined, the sill is the variance of the random process [or $\lim_{h \to \infty} \gamma(h)$].

Table D.1 Number of Observation Pairs versus Distance

Orientation	*h* (distance, m)	Number of pairs
North-south or east west, adjacent transects	1	12
Diagonally, adjacent transects	1.41	8
North-south or east west, skipping one transect	2	6
Diagonally, skipping two transects in one direction and one transect in the other	2.24	8
Diagonally, skipping one transect	2.83	2

In the preceding paragraph it is shown that a lag corresponds to a physical distance by virtue of the lag increment. By knowing the lag increment and the lag number we can determine the distance range for pairs of observations at a specific lag. The reader may ask "Why is it necessary to discuss lags when they are equivalent to distance?

The reason is that when sampling is not conducted using grid sampling or transects are not exactly perpendicular or parallel, the number of pairs of observations that will be separated by *a specific distance* will generally be only 1. By using a lag rather than a discrete distance, pairs of observations separated by a range of distances may be used to represent the covariance at the lag mid-point and a better estimate of the covariance is possible. Choosing larger lag increments results in a greater number of possible observation pairs but can obscure a variogram. Choosing lag increments is discussed in Subsection D.1.2.

A related idea that acknowledges the imprecision of an "exact" site location and also allows for a greater number of pairs to meet a distance criterion (but without increasing the lag increment) is a "lag tolerance." This is simply an allowable imprecision on a specific distance. Thus in the example described, with a lag tolerance of 0.01m, "bin 2" or "lag 2" would correspond to observations separated by a physical distance of 0.5 m - 0.01 m to 1.0 m + 0.01 m.

In this example, the number of observation pairs used to estimate the covariance at a specific distance varies with distance. In general, the larger the distance (or lag), the fewer the number of pairs of observations that meet the distance requirement for inclusion in that lag. One way to ensure that covariances are reasonably estimated is to specify the minimum number of observations that must be used to estimate the covariance. (Another way is to choose a larger lag increment).

Finally, because the degree of covariance among widely separated observations is generally less than closely spaced observations, the estimation of covariances beyond a given distance is of little utility. Thus a maximum lag is often specified when estimating covariances.

Most commonly employed software for estimating variograms provides default values for the lag tolerance, maximum lag and/or minimum number of observations used to estimate a covariance but the end-user should be aware of these concepts and ensure that the software defaults do not bias the choice of variogram model. Software inputs are discussed in further detail in Subsection D.1.2. Three commonly used variogram models are discussed below.

D.1.1 Three Commonly Employed Variograms

The spherical, exponential, and Gaussian theoretical variograms are commonly used when modelling[47] geostatistical processes. The most basic forms of these models omitting any nugget effect (non zero variogram at $h = 0$) are presented below. More complicated variants are available that include provision for anisotropy (Subsection D.1.5.1) and trends (Subsection D.1.5.2). Choosing between variograms is discussed in Subsection D.1.4; the remainder of this section discusses how to fit a variogram and what a variogram tells you.

The spherical variogram is

$$\lambda(h) = 1.5\left(\frac{h}{a}\right) - 0.5\left(\frac{h}{a}\right)^3 \text{ for } h \geq a; 1 \text{ otherwise} \qquad [20]$$

where:

- h is a given distance; and,
- a is the range.

The exponential variogram is

$$\lambda(h) = 1 - \exp\left(-\frac{3h}{a}\right) \qquad [21]$$

where:

- h is a given distance; and,
- a is the range.

The Gaussian variogram is

$$\lambda(h) = 1 - \exp\left(-\frac{3h^2}{a^2}\right) \qquad [22]$$

where:

- h is a given distance; and,
- a is the range.

The three theoretical variograms (using the same range) are plotted in Figure D.3.

D.1.2 Fitting the Empirical Variogram

The modelling process begins with fitting an empirical variogram using equation 19. In order to do so a computer program[48] is generally used. However the program will require some guidance in the same way that a person following a recipe needs more than a list of ingredients (the raw data and empirical variogram formula may be thought of as "ingredients") to create an "edible" end-product. A list of additional instructions follows. Note that many software programs contain default instructions and no additional instructions from the end user may be required. However the end user (as always) is responsible for the output generated using default instructions. The additional instructions required are:

1. The maximum lag (or distance, h) to use in estimating covariances. A common software default is ⅓ the maximum diagonal span of the spatial coverage although Journel and Huijbregts (1978) suggest using ½ this distance.

---

[47] A simple model is a formula that describes how the response (a biological test endpoint in this case) varies over space.

[48] Software used by the author includes ArcGIS Geostatistical Analyst, Surfer, S-plus Spatial Stats, various R-libraries including "spatial," "gstat,, "geoR."

2. The lag tolerance[49] which is the amount of allowable discrepancy between the desired lag and observed lags. This is necessary because the distances between points comprise a continuum rather than discrete observations. A common choice for the lag tolerance is ½ the lag increment.
3. The number of lags for which the covariance will be estimated.
4. A lag increment which is the distance between lags.

Note that only one of decisions 3 or 4 are necessary (because if the maximum distance is known but only the number of lags *or* the lag increment is known, the unknown (number of lags or lag increment) can be calculated). Some software may also require input regarding the minimum number of pairs for which a covariance is estimated. Journel and Huijbregts (1978) recommend a minimum of 30 pairs of observations to estimate a correlation (for a given distance or lag). Note that in the absence of this input it is the data analyst's responsibility to understand the limits of covariance for a given lag estimate from a few observation pairs.

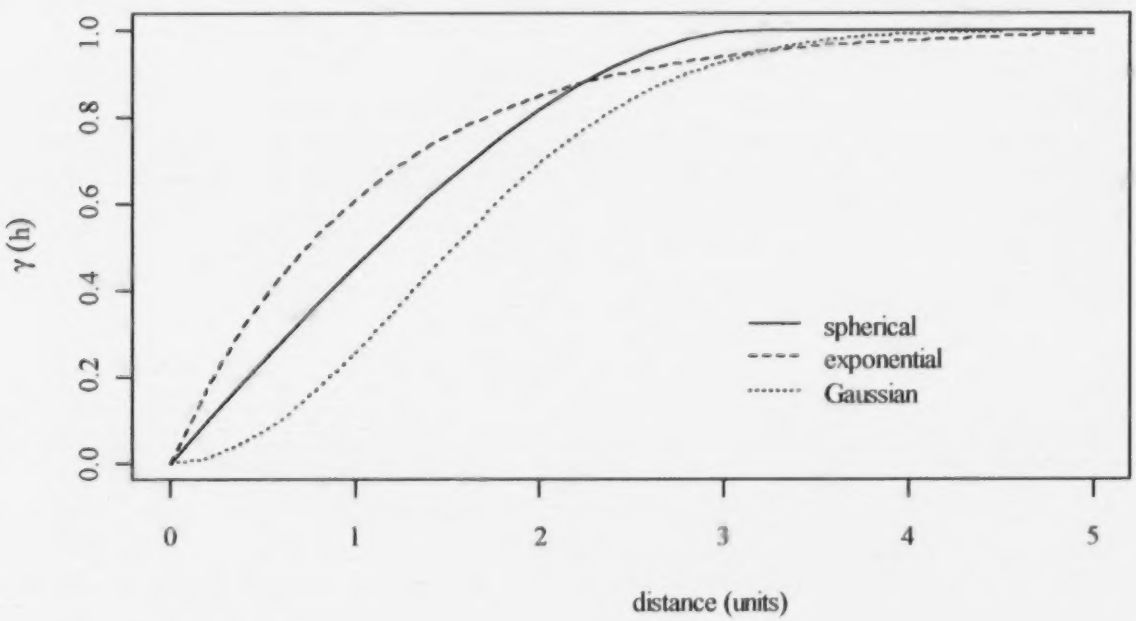


Figure D.3 Theoretical Variograms

D.1.3 What Does the Empirical Variogram Tell You?

The empirical variogram can tell you:

- Whether anisotropy (when the variance of observations is influenced by direction) or trends are present as suggested by a variogram without a sill; (these important concepts are discussed in Subsections D.1.5.1 and D.1.5.2, respectively);

[49] Lag tolerance is not part of the theoretical variogram. It is an operational construct used to inform the software what distances can be considered as the same. For example, if a lag is 50 m and a lag tolerance is 5 m any observations falling into the range 45 to 55 m would represent the 50 m lag.

- How correlated observations are within a given distance. This information is useful when conducting tests of hypotheses that assume observations are *not* correlated; estimating sample size to achieve desired error rates, etc.; and,

- The heterogeneity of the process. If the variogram at the origin (small distance increments or lags) is linear and/or the nugget is large, then the process (the biological responses over space) is irregular or non-smooth. If the variogram at the origin is parabolic then the process is regular.

D.1.4 Choosing a Variogram

Although there are objective criteria for fitting a variogram, the beginner geostatistical data analyst will often choose the best fitting variogram on a pragmatic basis. Fitting the variogram refers to estimating the parameters for the theoretical variograms, similar to estimating the parameters (slope, intercept, variance) of a regression model to fit a linear regression to x-y data. The variogram is fitted in order to have a description of how the correlation among observations varies with distance. This correlation model (fitted variogram) is used in the kriging process. Important considerations when fitting a variogram are to:

1. Fit the most obvious features of the empirical variogram. These are presence of a nugget and the presence/absence of a sill.
2. Use the simplest possible model that fits the most obvious features of the empirical variogram.
3. Use each candidate variogram to generate a kriged surface. Does the contour map describe the data well?[50] If all the candidate variograms describe the data well choose the simplest.

Depending upon the software used the following objective criteria may be available. The criteria may be used to compare contending models. However as mentioned earlier, it is important to note that choosing the variogram using these objective criteria requires predictions of the response being modelled at the sampling locations. The variogram describes the correlation structure among observations but by itself does not predict values. Thus prediction of some form is required to estimate the criteria used to choose a variogram. These criteria are placed in this section as a logical continuation of the discussion regarding variograms but practically, estimating the criteria follows modelling the observations themselves rather then correlation among observations. Examples of using these criteria to select a variogram are presented in Section D.2 (Appendix D).

The objective criteria listed below use residuals, which are the differences between the observed values and those predicted by the procedure used to model the observations. These criteria and the manner in which they are used are briefly described below; however, a more detailed discussion is beyond the scope of this document. Further details are presented in (Ripley, 1981; Isaaks and Srivastava, 1989; Cressie, 1993). The key point for the end-user is underlined. Note that it is unlikely that all of these criteria will be provided by a single software package; the end-user should use those criteria provided to determine which theoretical variogram is "best."

These criteria, although objective, are subject to limitations particularly if universal kriging is used for prediction.[51]

---

[50] The variogram is used (in a black box scenario) to generate a contour map of the biological responses. If the contour map disagrees with the measured responses it is a "bad" fit. For example if there is a biological hot spot where a biological response is very low and the contour map misses that low spot or indicates only a modest sag then the model is "poorly" fitting. The contour map is a "model" that describes the biological response, and that contour map relies on the variogram.

[51] The statistically inclined reader is referred to Cressie (1993) and the discussion on decomposition of small and large scale variation.

- **Mean prediction error** following validation.[52] This is the average of the residuals. *The average or mean prediction error should be "close"[53] to zero.*

- **Mean standardized prediction error** following validation.[52] This is the average of the standardized residuals. A residual is standardized by dividing it by a residual by the associated mean-squared prediction error. *The average or mean standardized prediction error should be "close" to zero.*

- **Root mean square prediction error** following validation.[52] This is the square root of the average of squared residuals. *Smaller is better.*

- **Root mean square standardized prediction error** following validation.[52] As above but uses standardized residuals. *Should be close to one.*

- **Median absolute deviation of validation residuals**. If the model predicts the dataset "well" the median absolute deviation of cross-validation residuals *will be small.*

- **Standard deviation of cross-validation residuals** is the usual estimate of standard deviation applied to the cross-validation residuals (these are defined above). *A smaller standard deviation is better than a larger standard deviation.* Note that the standard deviation is sensitive to outlying values.

- **Correlation between predictions and observed values**. *A high correlation between observed and predicted values is desirable.*

D.1.5 Variogram Requirements

The variograms described assume that the data are isotropic and non-trended. The lack of isotropy or anisotropy is discussed in Subsection D.1.5.1 and trends are discussed in Subsection D.1.5.2.

D.1.5.1 Anisotropy.

The variogram presented in Figure D.1 was generated using grid sampling as an example and distances, "*h*" were presented as scalars (single value) rather than as a set of vectors (distance and direction). The variogram is valid only if the variance between observations is a function *only of distance* and not direction. This is the assumption of isotropy. Note that this assumption is often not met but making the assumption is useful when introducing variograms.

If present, anisotropic behaviour in soil toxicity test responses will likely be due to gradients in contaminants and toxicity modifying factors (e.g., soil properties). That is the correlation among toxicity test responses will depend not only on the distances between observations but also on the direction between pairs of observations.

---

[52] Validation is the process of ensuring that the model predicts the observed values. The preferred approach is using data not used to fit the model. However such data are generally not available. Another approach to validation known as cross-validation is to use the fitted model and then remove a single observation from the dataset. Then, the fitted model is used to predict the deleted observation.

[53] "Close" and "small" are subjective adjectives and should be judged on the basis of experience or contextually. A contextual evaluation is possible by examining a criterion estimated for a set of theoretical variograms. The variogram generating a criterion "closest" indicated value (zero in the case of mean prediction error) be used to choose the "best" variogram; in terms of the criterion being evaluated. Similar concepts apply to the adjective "small." Generally, the different criteria will all indicate the same "best" variogram.

It is necessary to identify anisotropy when present because the variograms used for kriging rely on isotropic covariance models.

Identifying Anisotropy. Anisotropy may be identified in two ways. In the first method a generally increasing variogram (a variogram with no sill) indicates anisotropy. The second method uses directional variograms. A directional variogram places directional restrictions on the pairs of observations that may be used to generate a variogram. For example the angle relative to north or azimuth may be set at $\theta = 45°$ with a directional tolerance of 45°. Thus pairs of observations are restricted to those:

- separated by the distance or lag *"h"* ± the lag tolerance; and,
- along the direction 45° ± the azimuth tolerance.

If the variograms along different azimuths differ from one another the data are anisotropic. In this case a statistician should be consulted.

D.1.5.2 Trends.

The variogram describes how observations separated by a distance are correlated. One of the assumptions of a variogram is that the mean is constant. If there is a trend in observations the mean is not constant.

A trend in observations may be identified

- graphically using a 3-dimensional graphic showing the biological response as a function of the "x" and "y" site coordinates; these may be latitudes and longitudes or incremental distances from a fixed initial point; or
- by an empirical variogram without a sill (although note that this may also indicate anisotropy).

Two approaches to dealing with trended geostatistical data are to remove the trend or use a method that incorporates the trend. An example using the latter method is provided in the case study.

The reader is directed to Subsection D.2 which provides a description of how to apply these geostatistical concepts using real and synthetic datasets.

D.2 Geostatistics Case Studies Using Real and Synthetic Datasets

The material provided in this subsection is presented as three case studies to provide step-by-step guidance on how to use geostatistical methods with contaminated site data to generate stochastic contour maps of observed biological responses and/or chemical observations. These case studies compliment the guidance provided in Subsections D.1 and 3.3.6, and are intended to serve as a tutorial of sorts.

The following case study uses three datasets in order to demonstrate a complete analysis.

1. The first dataset consists of soil sample composites collected using a systematic grid sampling strategy comprised of 15- × 15-m plots. The plots are found south and east of the river Meuse near the village of Stein. The Meuse River begins in France and drains into the North Sea passing through Belgium and The Netherlands. The 155 soil samples were analyzed for heavy metals (Cd, Cu, Pb, and Zn). The data are available in Tables D.3, D.4, and D.5. Although the measurements are soil metal concentrations the methods describe apply equally to soil toxicity test measurement endpoints. These data have been modified for the purposes of this case study.

2. The second dataset is a synthetic dataset comprised of 421 observations simulated using a spherical covariance function (a spherical variogram). The points are simulated on a square grid of 21 × 21 locations. This is an example where aside from correlation among adjacent observations the toxicity test responses are random over the area being studied (there is no trend over the site to the biological test responses).

3. The third dataset uses the second dataset and superimposes a trended surface. In this case there is a clear gradient in the toxicity test response over the area being studied.

The purpose of the assessment is to produce a "map" of the response (soil Zn or biological test response) at unmeasured locations with prediction errors. The process of generating a map is presented below as a set of steps but the process is not as unidirectional as it appears; some iteration is often required. Also, note that some software may automate one or more of these steps.

1) Data quality assurance is critical. This step consists of estimating and examining summary statistics to detect unusual observations. Any unusual observations should be confirmed with the biological testing laboratory prior to proceeding with analysis. Additionally the data should be plotted and examined visually. Look for trends (Subsection D.1.5.2; Appendix D).
2) Generate the empirical variogram and estimate "by eye" the nugget, sill and range. Some computer programs require initial estimates as starting points; the values will also be used to detect very obvious mistakes.
3) Look for anisotropy (Subsection D.1.5.1; Appendix D).
4) Fit one or more theoretical variograms that might describe the empirical variogram.
5) Examine the fitted variogram(s) for any obvious inconsistencies with the empirical variogram.
6) Fit the kriged model using the selected variogram from the previous step.
7) Examine objective criteria provided by the software *and* examine the fitted surface. If necessary, choose a different theoretical variogram and repeat.
8) Generate the prediction standard error plot.

Step 1 — Data Quality Assurance

This tutorial assumes that all the data have been quality assured numerically. Only graphics are presented below and three different types of graphics are presented to show some of the possibilities. The same graphic type is maintained for a given dataset. Each graphic is examined for patterns in concentration as well as anomalous results.

Practical Tip: The reader should be aware of the aspect ratio (distance that a vertical distance represents relative to a horizontal distance) used to generate a graphic. An aspect ratio of 1 presents the responses "as is" whereas aspect ratios (for the datasets being considered) other than 1 will distort the visualization.

Examples of Methods of Graphically Presenting Data

The graphic for dataset #1 (Figure D.4) uses a "bubble" plot with a symbol at the sample location. The size of each symbol is proportional to the magnitude of the response (modified soil log (Zn) concentrations) and the aspect ratio is 1.

The graphic for dataset #2 (Figure D.5) uses symbols of the same size but uses colour and a legend to distinguish between ranges in the continuous response instead of size as the bubble plot does. This plot uses an aspect ratio of 1.

The graphic for dataset #3 (Figure D.6) is a contour plot with an aspect ratio $\neq 1$. Contour plots are useful when the data surface is convoluted.

Practical Tip: Readers should be aware that *any* contour plot uses a default mathematical algorithm to generate the contours. The choice of algorithm and/or input variables will affect the plot produced and can bias interpretations thereof. The user should generate several graphics modifying the default algorithm to ensure that conclusions are not biased by the algorithm used.

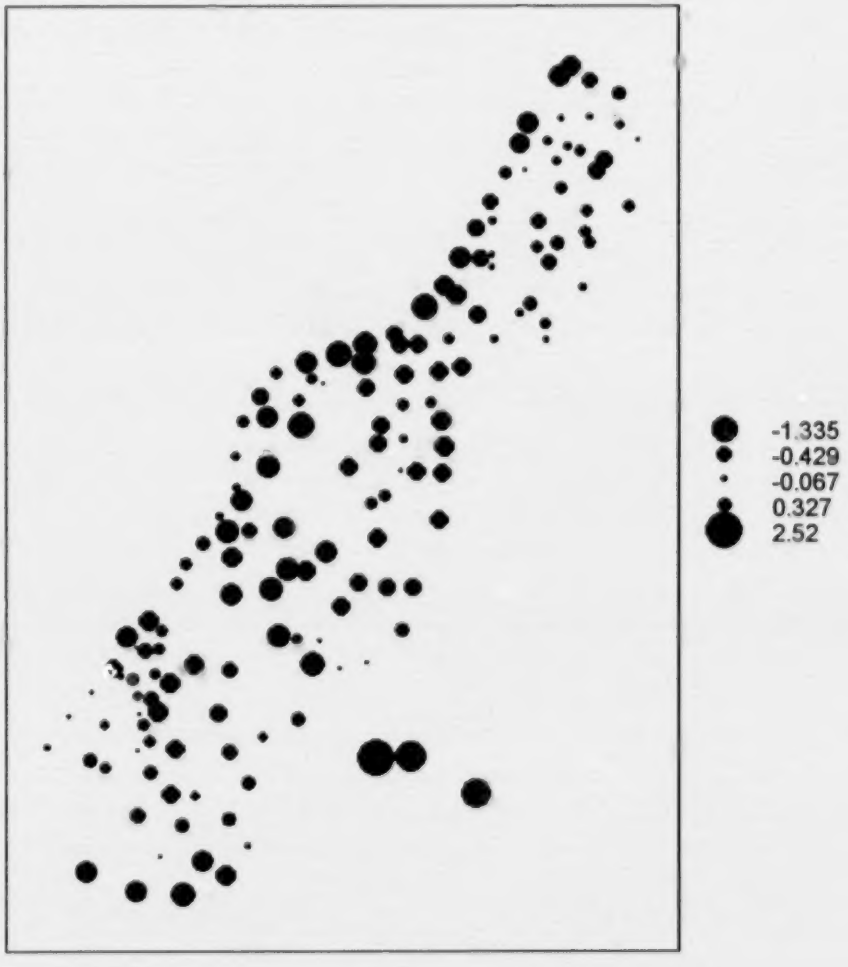


Figure D.4 Dataset #1: Spatial Plot of Concentrations

No obvious trends in the log (Zn) concentrations are apparent although possibly higher concentrations appear along the shoreline (to the north and west) than might if the process responsible for the Zn concentrations was truly random.

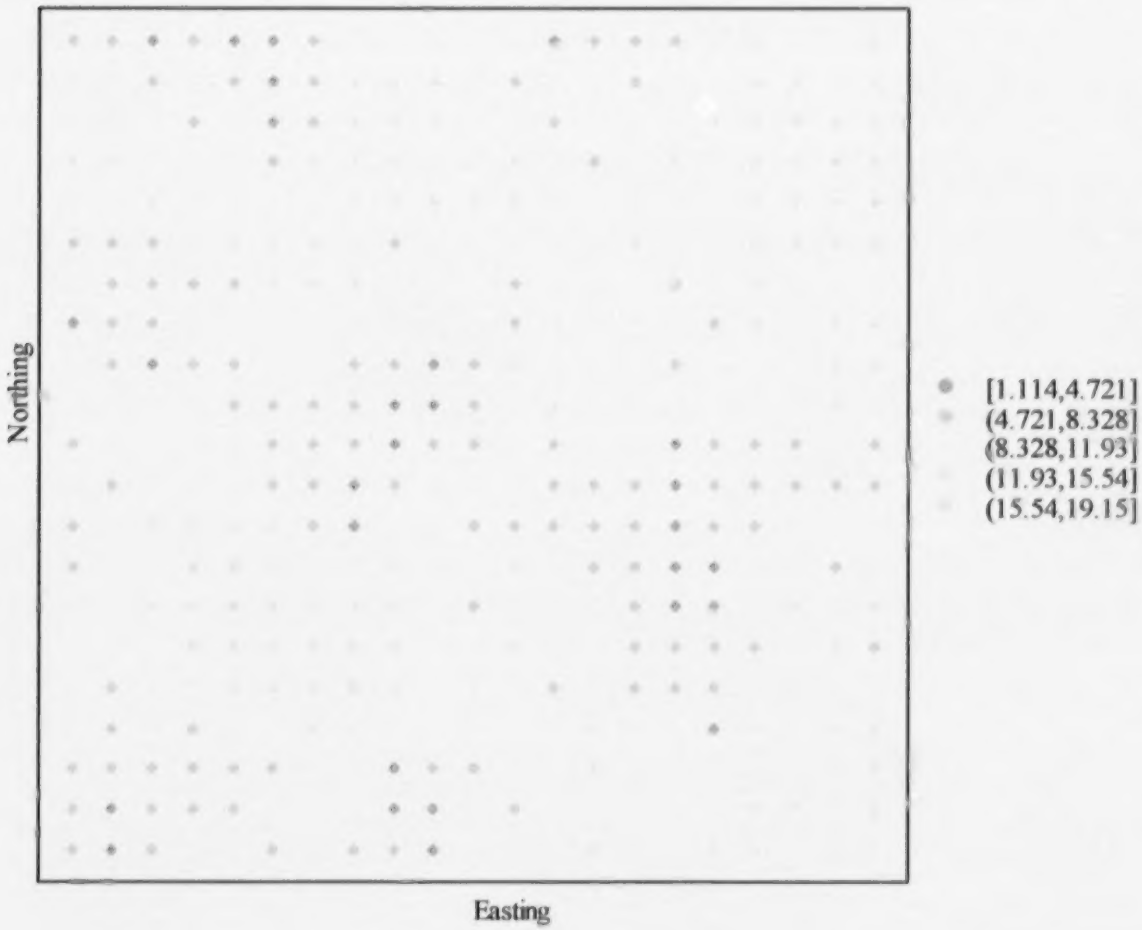


Figure D.5 Dataset #2: Colour and Legend Plot

The figure above for the synthetic dataset shows there are elevated patches that are seemingly random. The numeric ranges following a colour in the legend indicate the magnitude of the simulated biological response.

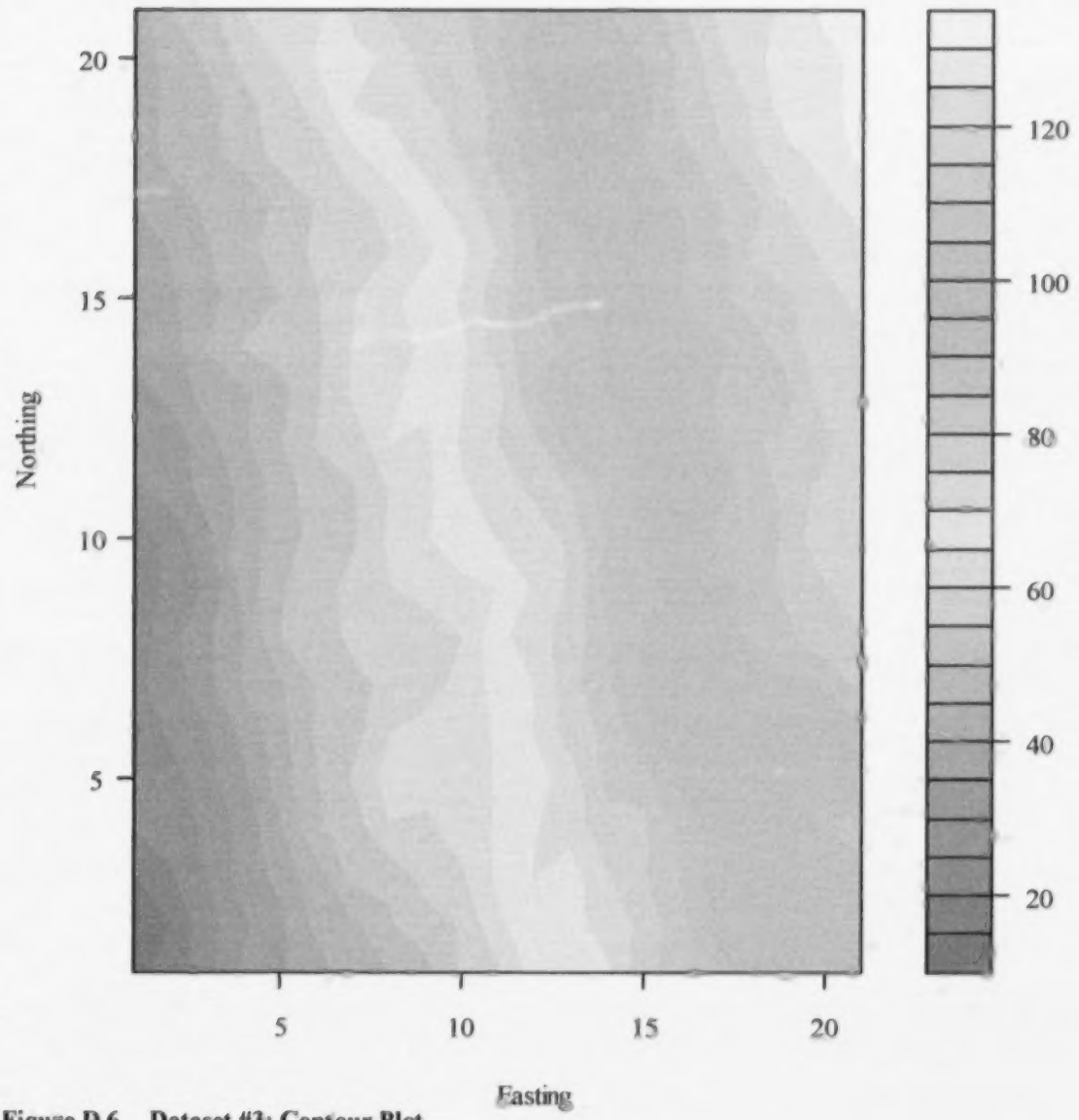


Figure D.6 Dataset #3: Contour Plot

The figure above for the synthetic dataset with a trend shows the biological response increases in a roughly north-northeast direction. Recall that the aspect ratio ≠ 1 and therefore the northeast is *not* represented by a 45° line on this graphic.

A Note on Default Settings

Software usually[54] provides reasonable default settings necessary to fit empirical variograms (distance or lag tolerances, azimuth tolerances, minimum number of data pairs to estimate a semivariance, etc.), kriged

[54] Software used by the author includes ArcGIS Geostatistical Analyst, Surfer, S-plus Spatial Stats, various R-libraries including "spatial," "gstat," "geoR."

surfaces (minimum and maximum number of neighbours, etc.). The reader should begin data analysis using the default settings.

Step 2 — Fitting the Empirical Variogram

The empirical variogram for dataset #1 (the soil Zn dataset) is presented in Figure D.7.

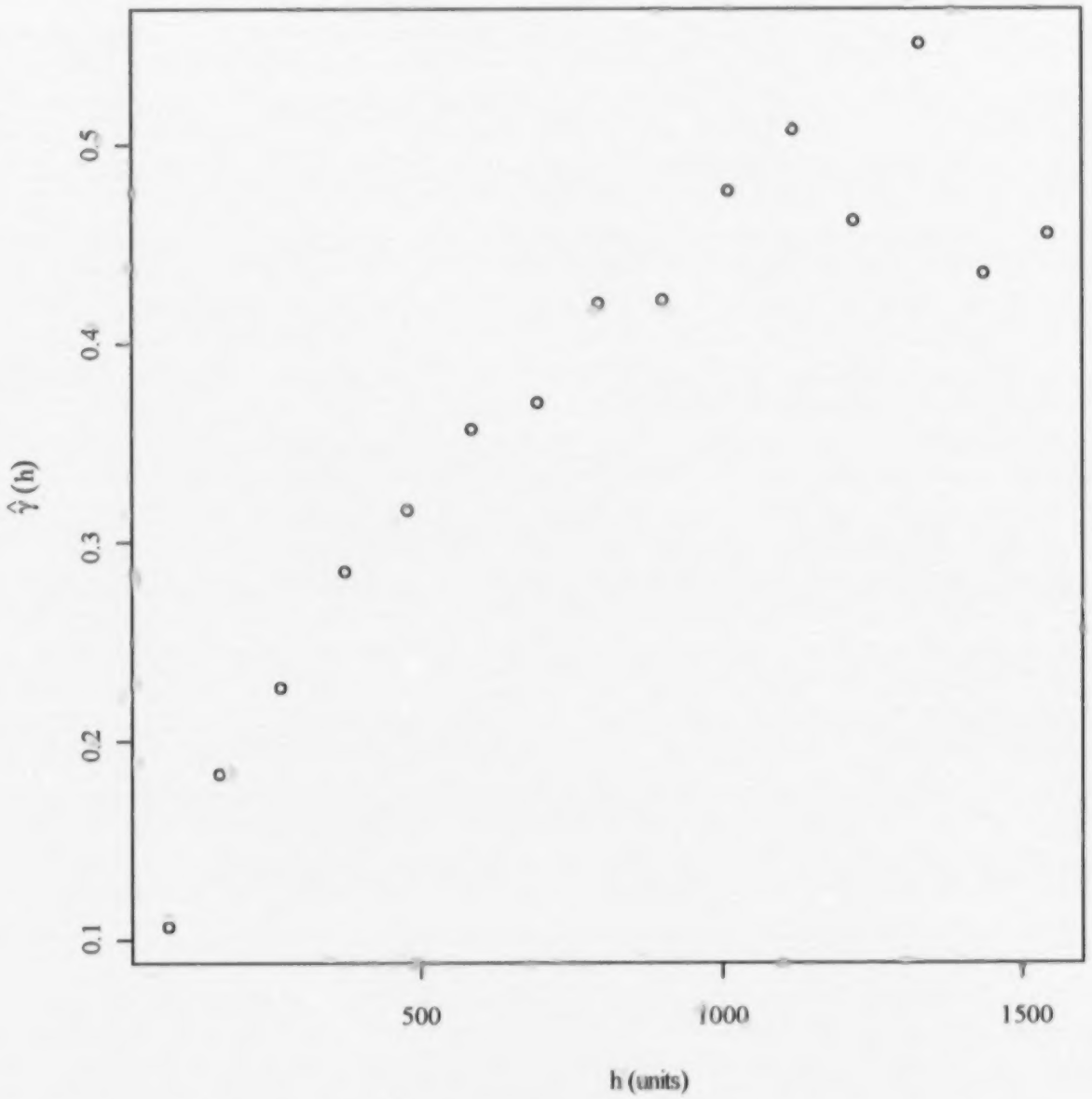


Figure D.7 Empirical Variogram, Dataset #1 (Soil Zn Data)

Depending upon the software used, initial estimates of the sill, nugget and range may be necessary. These may be obtained from the empirical variogram above. The empirical variogram above rises from 0.1 which is the micro-scale variation or nugget. The variogram seems to flatten at $\hat{\gamma}(h) = 0.48$ although it might

also increase and not flatten out. Assuming that the variogram does flatten (there is a horizontal asymptote), the range might be 1000 units.

The empirical variogram for dataset #2 (the synthetic data) is presented in figure D.8.

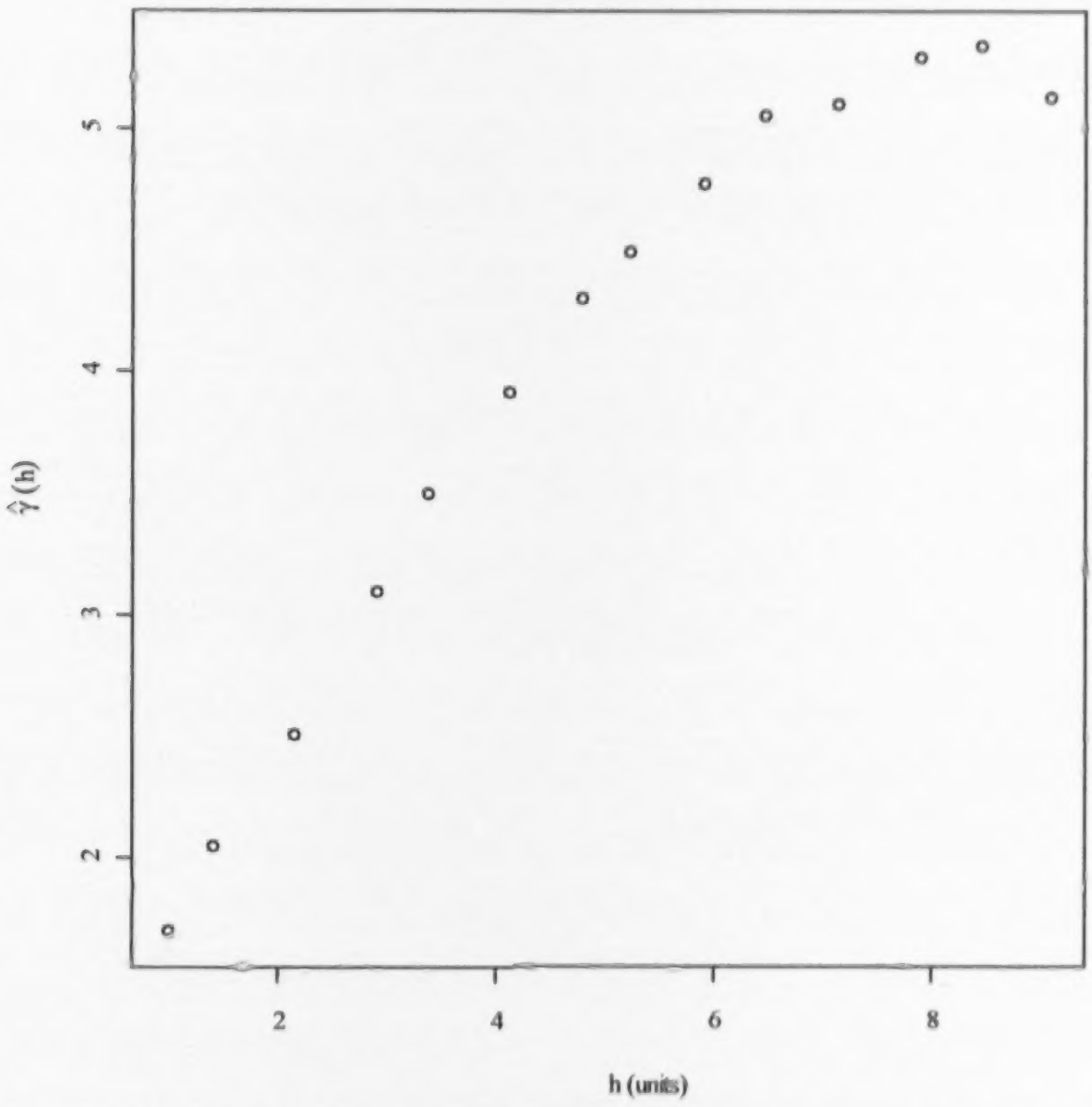


Figure D.8 Empirical Variogram, Dataset #2 (Synthetic Data)

The empirical variogram above rises from 1.5 which is the micro-scale variation or nugget. The variogram seems to flatten at $\hat{\gamma}(h) = 5.2$ and the range appears to be 8 units.

The empirical variogram for the dataset #3 (trended synthetic data) is presented in Figure D.9.

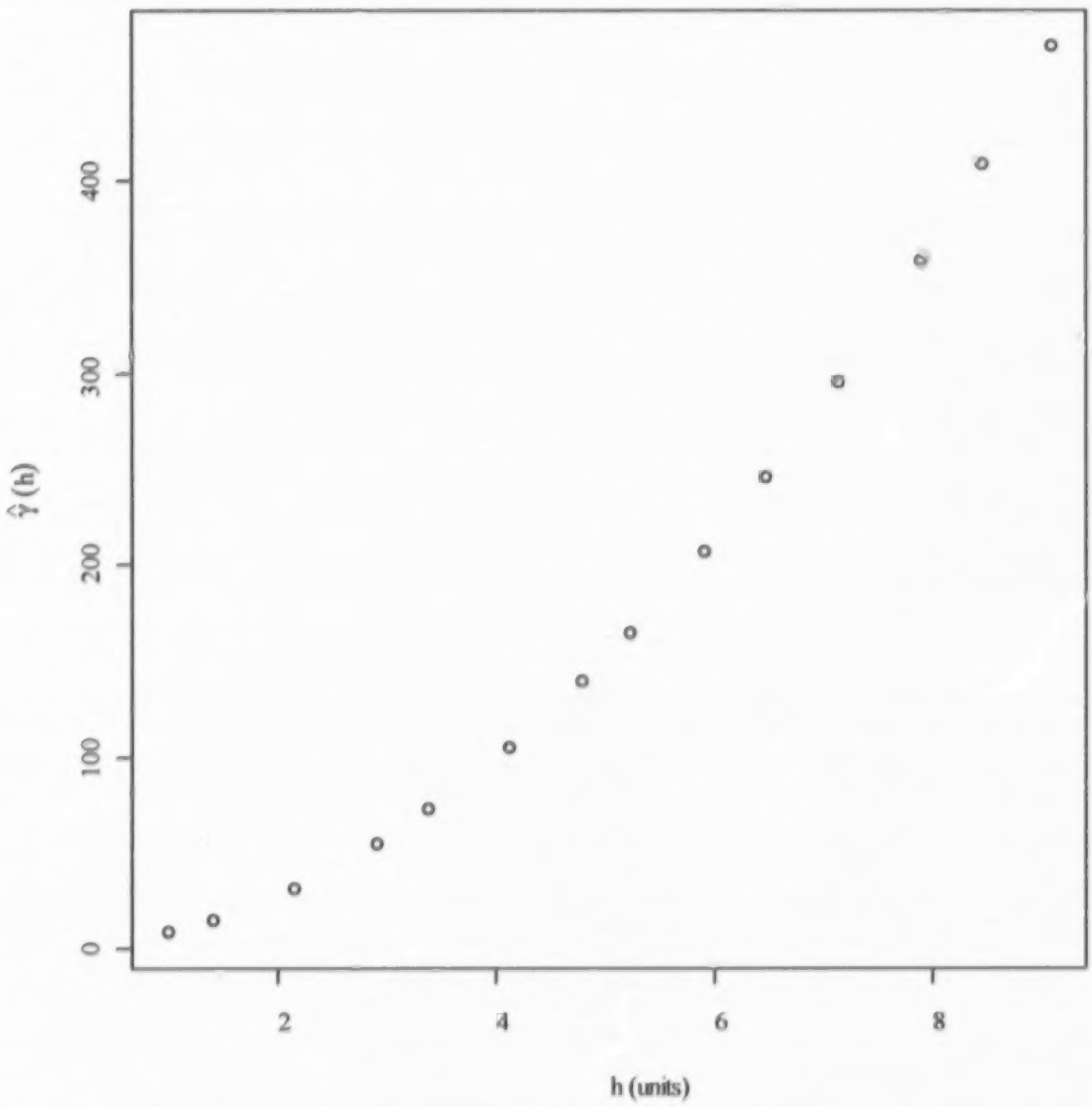


Figure D.9 Empirical Variogram, Dataset #3 (Synthetic Dataset with Trend)

The empirical variogram above shows no evidence of a plateau (sill), which is expected, as the data are visibly trended (Figure D.6).

Dealing with Trends

The empirical variogram for synthetic trended dataset (dataset #3) shows no evidence of a sill, which is expected, as the data are visibly trended (see Figure D.6). It is necessary to remove this trend prior to fitting variograms or kriging. This process is known as detrending. The detrending process involves modelling the relationship between the response and the independent variables and using the residuals from the fitted model as input to the kriging or variogram algorithm.

As is typical of statistical modelling in general, a model to describe a response depends upon the statistical distribution of the response being modelled and the supposed relationship between the response and the explanatory variables. For example a survival response that shows a curvilinear pattern over an area might best be modelled using polynomial logistic regression. A continuous response that with a simple gradient over an area might be modelled using a familiar linear regression model with the "x" and "y" coordinates as independent variables. Guidance on detrending in general is beyond the scope of this document; the example selected will often (but not always!) be appropriate. Guidance from a statistician may be necessary to detrend some responses.

Operationally, detrending may be accomplished within a variogram or kriging algorithm[55] (the software used specifically to fit a variogram or kriged surface) *or* prior to submitting data to the variogram or kriging algorithm. The latter method is chosen here to better illustrate the process.

The "response" in dataset #3 (trended synthetic dataset) is continuous with a range of approximately 10 to 135 units (numbers are derived from dataset #2). The response might represent something such as plant root lengths reported in mm. The explanatory variables are (in this simple example) the "x" and "y" locations. This suggests that a regression model under the assumption of normality might be appropriate. However, the structure apparent in the empirical variogram (and the fact that in this synthetic dataset a correlation structure was built in) suggests that assumption of independence of observations required for ordinary least square regression will not be met. Given that the trend model will be used only to detrend the surface as a preparation for fitting a variogram and kriging, ordinary least squares regression is used. If this were a real scenario, one possible option would be to use generalized least squares.

The following model was fit to the trended synthetic data:

$$\text{Response} = 9.317 + 1.385x + 4.540y.$$

Details of the fitted linear regression model are not presented as they are ancillary to detrending (but still critical in that the model does describe the trend). Using this model the residuals (difference between observed and predicted values) are obtained and used as input to the kriging and variogram algorithms.

After de-trending the empirical variogram is Figure D.10:

[55] This is known as universal kriging or less commonly as external drift kriging if the independent variables do not include coordinates.

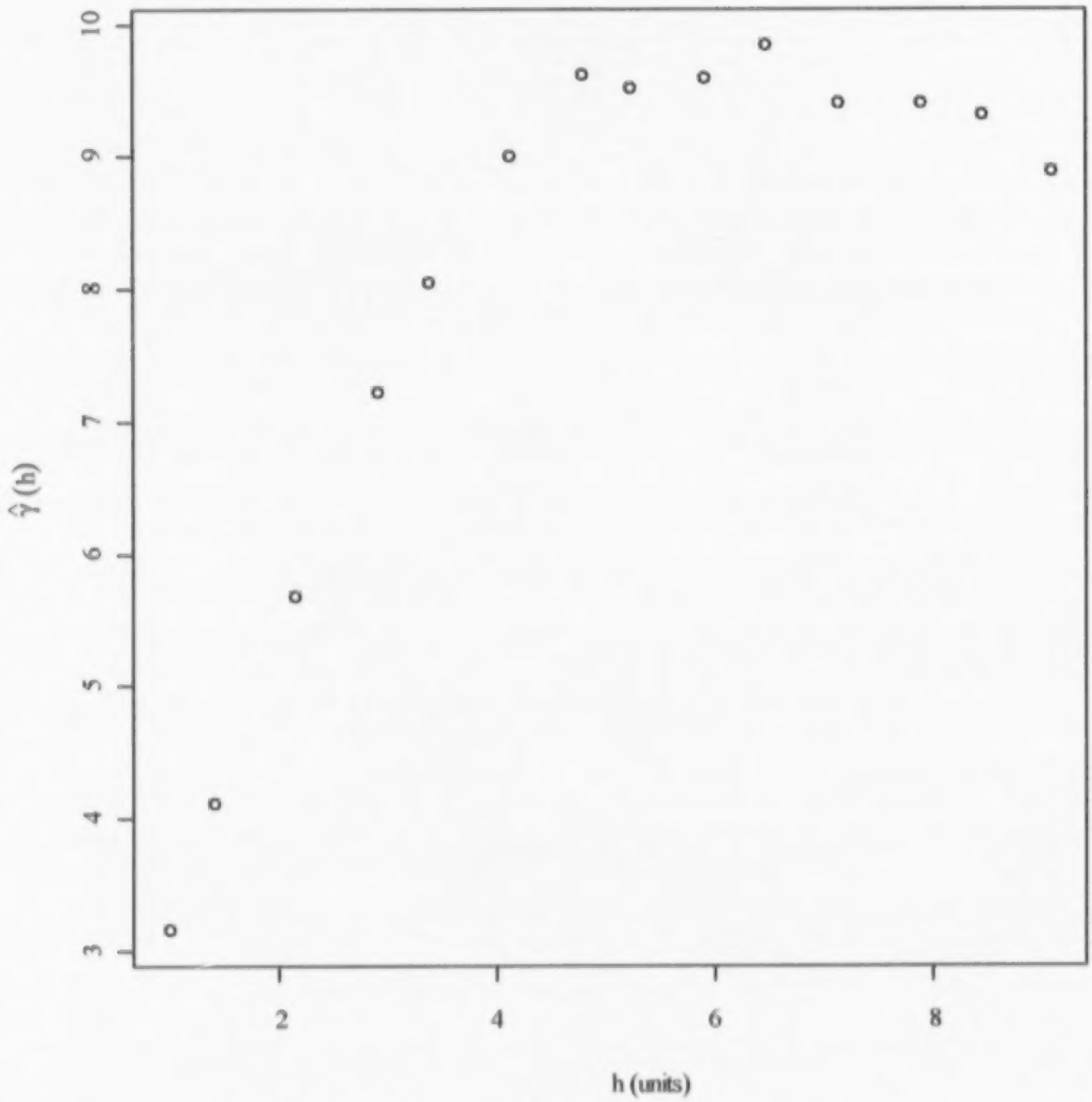


Figure D.10 Empirical Variogram, Dataset #3 (Detrended Synthetic Data)

The empirical variogram of the detrended data[56] now has a plateau or sill of approximately 9.5, a nugget of 3 and range approximately 5 units.

[56] Note that the process of superimposing a trend over dataset # 2 also adds some random variation around the trended surface. Thus the empirical variogram for the detrended data presented in Figure D.10 is not identical to the variogram for the original data (dataset # 2).

Step 3 — Look for Anisotropy

Anisotropy may be identified using directional variograms. These can be produced in separate plots although some software (such as ArcGIS Geostatistical Analyst) produces directional variograms in "real-time" by rotating an azimuth. Variograms for azimuths of 0, 45, 90, and 135° are presented in Figure D.11. Note that it is unnecessary to create additional directional variograms for azimuths between 180 and 360° due to the symmetry of $\gamma(h)$.

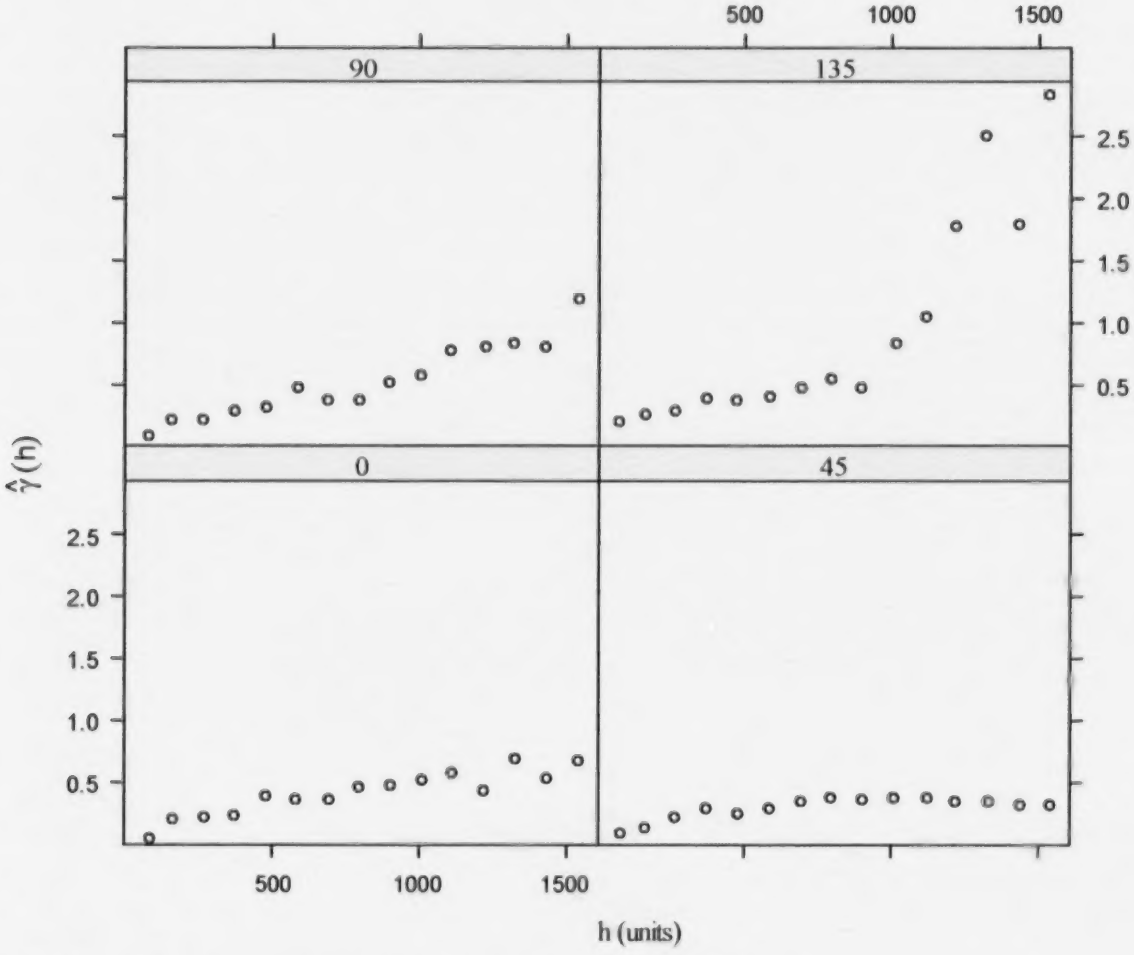


Figure D.11 Directional Empirical Variograms, Dataset #1 (Soil Zn Data)

The directional empirical variograms for the soil Zn dataset vary considerably based on direction. As this point a statistician should be consulted because the assumption of isotropy necessary for kriging is clearly not met. Therefore no further examination of this dataset will be considered in this case study.

The case study continues with the second dataset (synthetic data) presenting the empirical variogram and directional variograms (see Figure D.12).

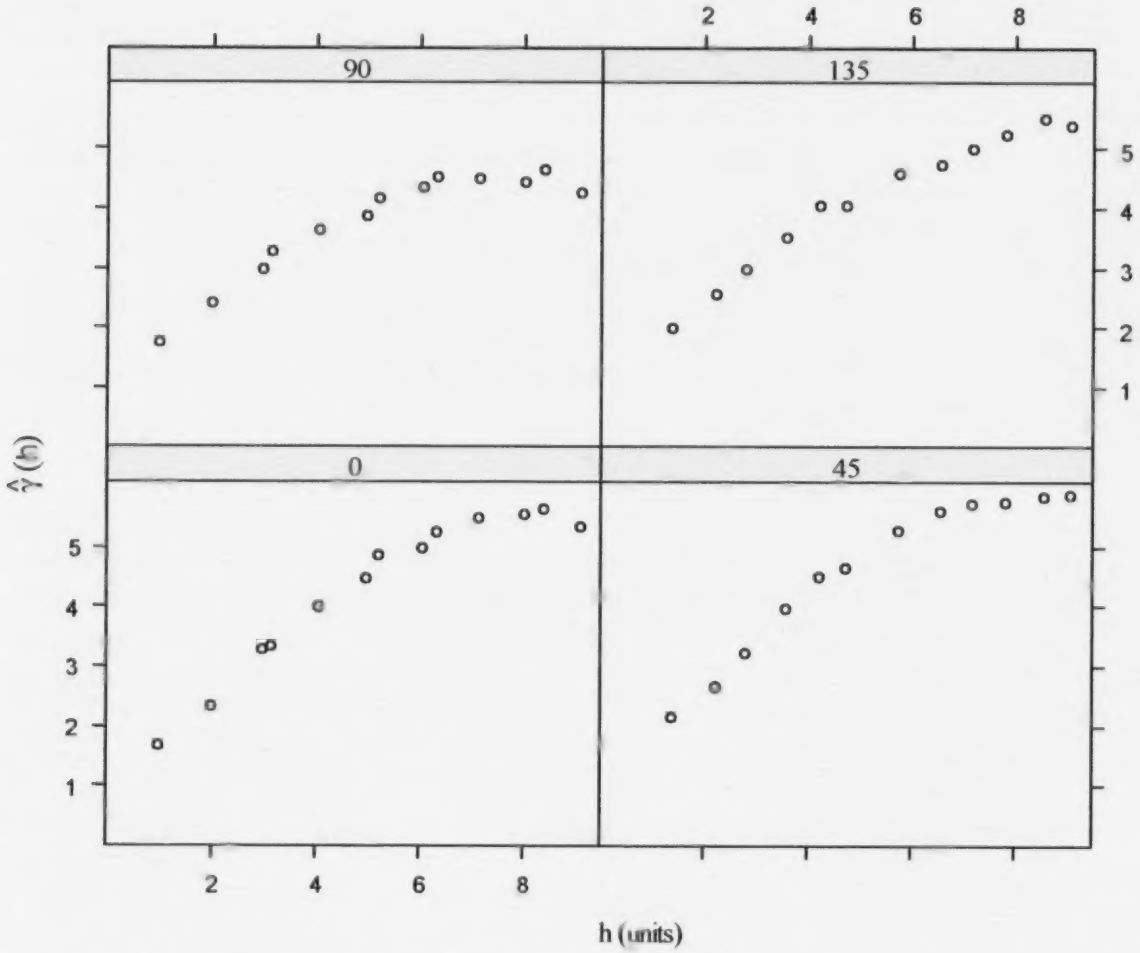


Figure D.12 Directional Empirical Variograms, Dataset #2 Synthetic Data

The directional empirical variograms for the synthetic data above all appear similar in terms of nugget, sill, and range. The assumption of isotropy appears to be satisfied.

No directional variograms are estimated for the synthetic trended data as the data are obviously trended and therefore anisotropic.

The directional empirical variograms for dataset #3 (de-trended synthetic data) are shown in Figure D.13.

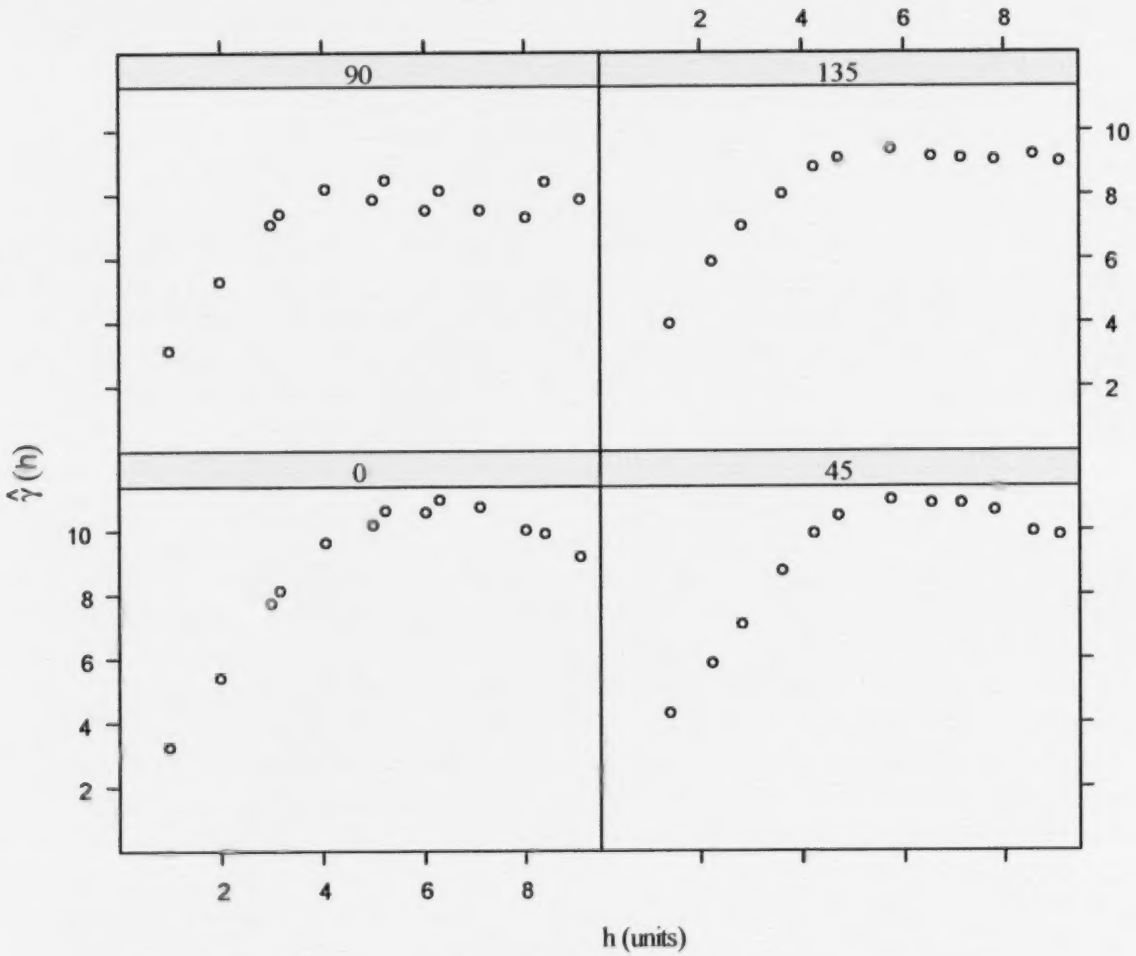


Figure D.13 Directional Empirical Variograms, Dataset #3 (De-trended Synthetic Data)

The directional empirical variograms for the de-trended synthetic data above all appear similar in terms of nugget, sill and range. The assumption of isotropy appears to be satisfied.

Steps 4 and 5 — Fit Theoretical Variograms

Empirical variograms were fitted for both datasets and were used to generate initial estimates of the nugget, sill and range. Depending upon the software used the initial estimates might be required to fit a theoretical variogram (Figure D.14).

A Practical Tip: If the algorithm fitting the variograms displays a message such as "failure to converge" the initial parameter estimates may be erroneous or insufficiently close to the true values. Several combinations of initial values should be tried.

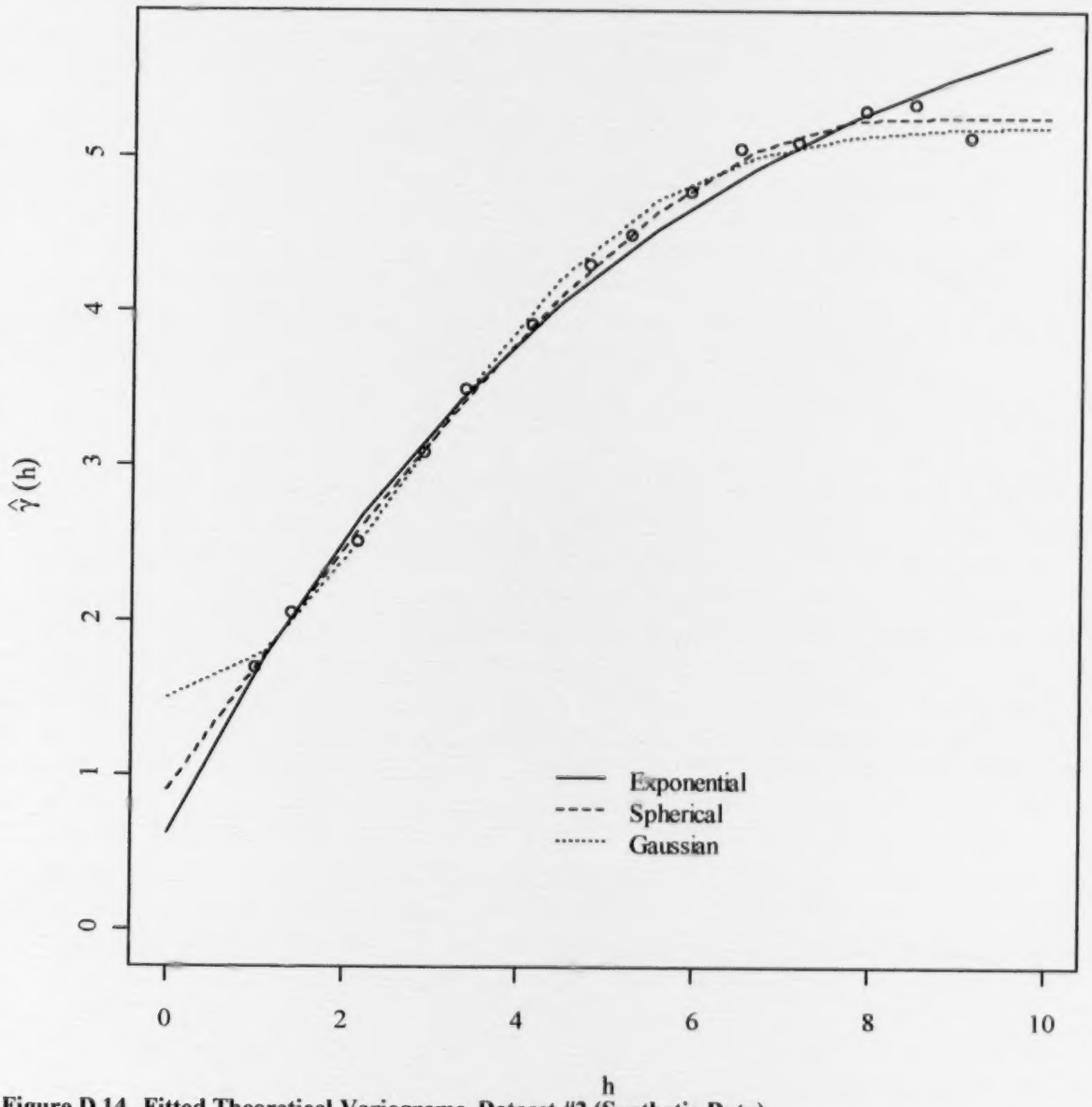


Figure D.14 Fitted Theoretical Variograms, Dataset #2 (Synthetic Data)

The fitted theoretical variogram above suggests that a spherical covariance structure (spherical variogram) best describes this dataset. This is gratifying as the data were simulated using this covariance structure!

Steps 6 and 7 — Fit and Use the Kriged Model (using Dataset #2, Synthetic Data)

The software used to conduct kriging will produce a variety of output. This includes a map of the fitted responses, prediction standard errors and some objective criteria that may be used to choose among various possible models. In the simple case described here where the biological measurement responses vary only as a function of "x" and "y" coordinates the only possible user input to model structure is the choice of variogram model.

Although the simulated data used a spherical covariance structure (spherical variogram), at this point we pretend that this is not known and fit the kriged surface using ordinary kriging, three times, assuming a spherical, exponential or Gaussian covariance structure (theoretical variograms, refer to subsection D.1.1; Appendix D).

Table D.2 Criteria for Choosing "Best" Kriging Model

	Criteria				
Theoretical Variogram	Mean prediction error	root mean prediction error	mean standardized prediction error	root mean standardized prediction error	median absolute deviation
exponential	-0.0004276	1.716578847	-0.000142441	0.975897027	1.190025
spherical	0.00078189	1.718307818	0.000191322	1.032990888	1.1748353
Gaussian	-0.0001409	1.761149643	-4.95E-05	0.917285562	1.2196648
What is "desirable"	close to zero	smaller is better	close to zero	close to one	smaller is better

Given that in this simple example the only user-input to model structure is the choice of theoretical variograms the table might really be entitled "criteria for choosing best variogram." Keeping the results of this table and Figure D.14 in mind it is clear that the Gaussian variogram does not describe the covariance structure of the data as well as the spherical and exponential variograms. The objective criteria may suggest that the spherical variogram performs better than the exponential variogram; Figure D.14 confirms this.

Step 8 – Generate the Prediction Standard Error Plot

Using the kriging model the biological response can be "predicted" at unmeasured locations. However it is important to note that predicted values cannot predict something that has not been measured. Thus toxicity hot spots that occur on a scale finer than the sampling scale may be missed, notwithstanding that a kriged surface might state otherwise.

The following graphic (Figure D.15), using a much finer scale than that measured, is generated using ordinary kriging under the assumption of spherical covariance.

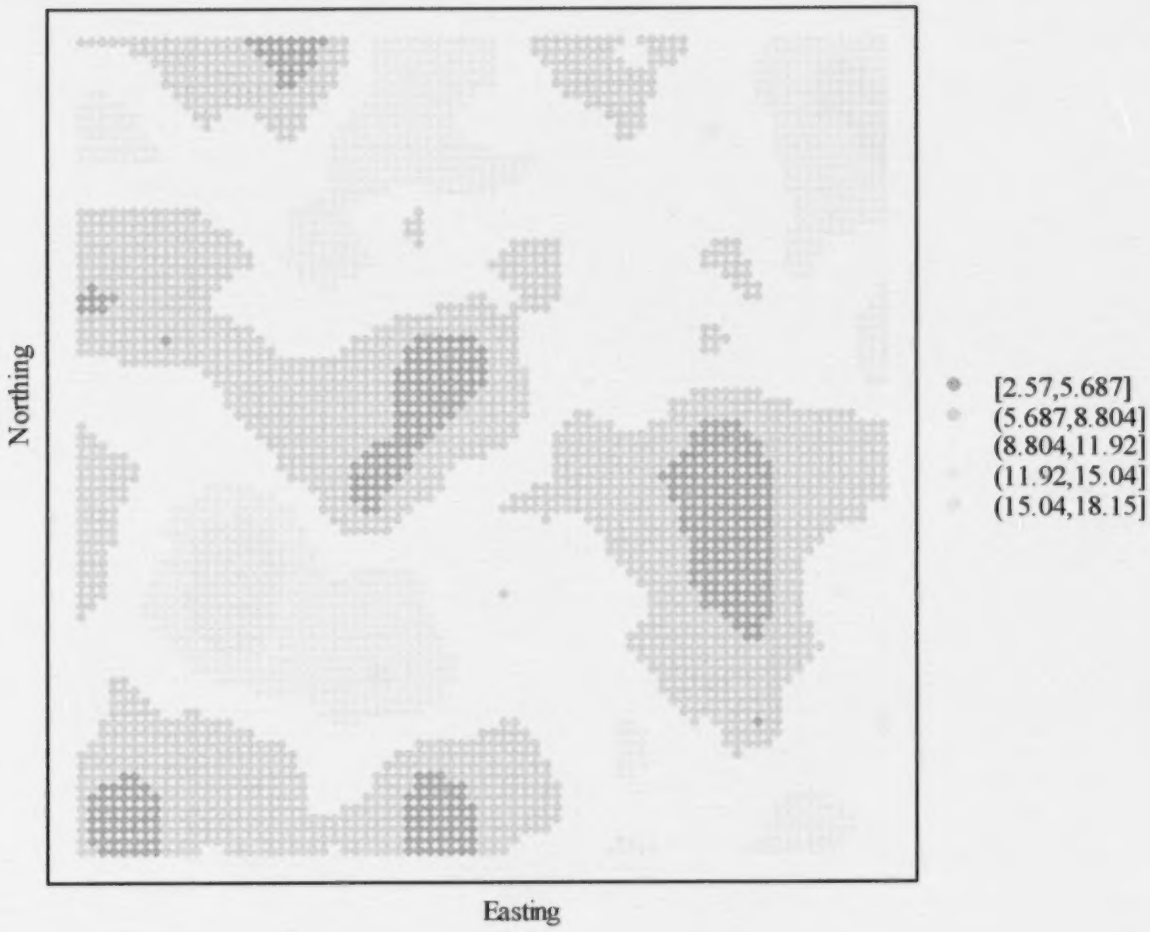


Figure D.15 Kriged Surface, Dataset #2 (Synthetic Data)

The kriged surface mimics the information presented in the plot of the raw data and allows for prediction at locations not sampled (although see caveat above). This simple kriged surface may be used to:

1. predict a response at a specific location;
2. examine areas where the variance for a predicted value may be unacceptably high (see Figure D.16); and,
3. delineate areas bounded by an unacceptable response.

A response may be predicted at a location where no sample was collected using the fitted kriged model. The most practical way to do this is to use the modelling software and appropriate option to predict

observations at unsampled locations which are accompanied[57] with standard errors on the predictions. Assuming normality a confidence interval around the predicted value can be generated with the usual formula.

Prediction errors over sub-areas can also be assessed visually using graphics. The following plot of prediction standard errors (Figure D.16) reverts back to the soil Zn (dataset #1) example because the synthetic dataset imparts regularity to the kriged prediction standard error surface that is of little interest.

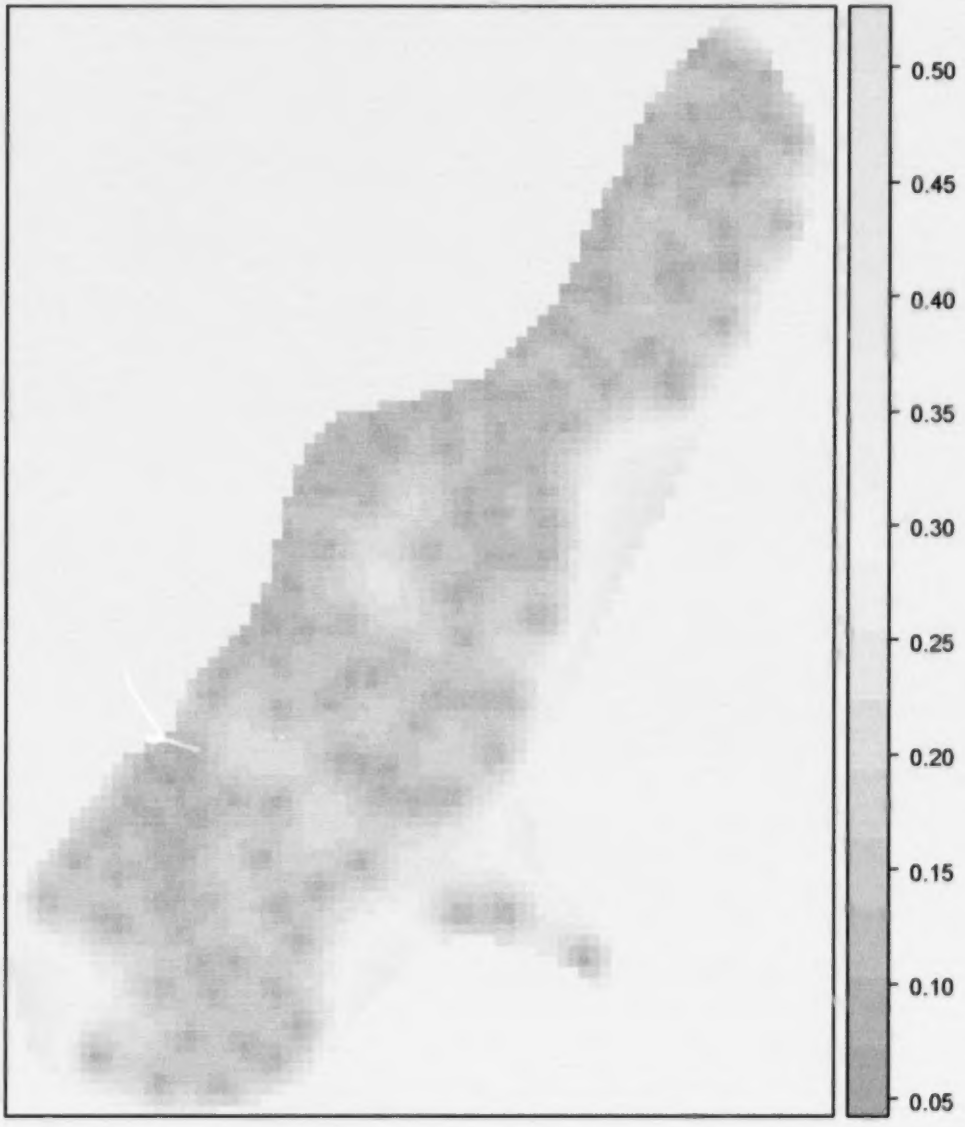


Figure D.16 Kriging Prediction Standard Errors, Soil Zn Data (Dataset #1)

[57] This statement is based on experience with the following software: ArcGIS Geostatistical Analyst, Surfer, S-plus Spatial Stats, various R-libraries including "spatial," "gstat," "geoR. "

Figure D.16 shows that the lowest standard error is 0.05, while the largest is greater than 0.40, so there is as much as an 8-fold change in prediction standard errors over the area in soil Zn was measured. This has implications with respect to decision-making; some areas in the study area are better predicted than others. If decision-making is predicated upon responses in the less precisely predicted areas, additional sampling may be warranted. The reader should refer to Subsection 3.3.5.1 (Error Rate Control) for a discussion on precision and decision errors.

Figure D.17 shows how a fitted model can be used to delineate areas. The example uses the trended spherical data (dataset #3). Colours are used to code biological responses as follows:

- < 25 "unacceptable" (red);
- (25,50) "of concern" (orange);
- (50,70) " of minor concern" (yellow); and
- > 70 "of no concern" (green).

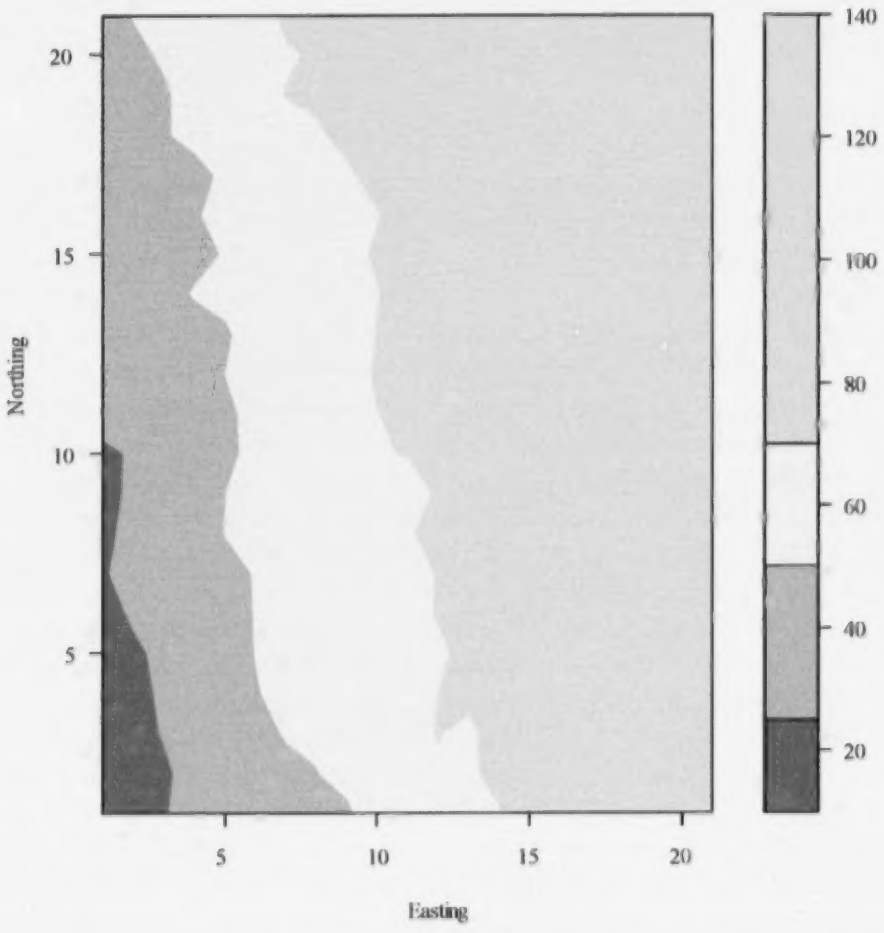


Figure D.17 Delineating Isopleths on a Kriged Surface to Inform Decision-making (Dataset #3)

Using the criteria for the biological response provided, Figure D.17 suggests only a small proportion of the area to the southwest exhibits an unacceptable response.

Table D.3 Dataset 1

x	x-coordinate
y	y-coordinate
Zn	Zn concentration used in kriging
lzinc.res	modified Zn concentration used to generate Figures D.4, D.7, and D.11

x	y	Zn	lzinc.res	x	y	Zn	lzinc.res
181072	333611	1022	0.748343	180561	332193	167	-0.61027
181025	333558	1141	0.849048	180451	332175	176	-0.64985
181165	333537	640	0.417017	180410	332031	258	-0.21109
181298	333484	257	-0.33471	180355	332299	746	0.621832
181307	333330	269	-0.17894	180292	332157	746	0.65601
181390	333260	281	-0.01066	180283	332014	464	0.267037
181165	333370	346	-0.0878	180282	331861	365	0.127077
181027	333363	406	-0.05368	180270	331707	282	-0.04062
181060	333231	347	-0.09241	180199	331591	375	0.253857
181232	333168	183	-0.52813	180135	331552	222	-0.30513
181191	333115	189	-0.49963	180237	332351	812	0.560778
181032	333031	251	-0.31075	180103	332297	1548	1.115001
180874	333339	1096	0.810642	179973	332255	1839	1.192125
180969	333252	504	0.180988	179826	332217	1528	0.893024
181011	333161	326	-0.15494	179687	332161	933	0.305314
180830	333246	1032	0.770268	179792	332035	432	-0.28228
180763	333104	606	0.268292	179902	332113	550	0.011684
180694	332972	711	0.449992	180100	332213	1571	1.182351
180625	332847	735	0.500478	179604	332059	1190	0.537496
180555	332707	1052	0.885298	179526	331936	907	0.273393
180642	332708	673	0.520151	179495	331770	761	0.178141
180704	332717	402	0.05751	179489	331633	659	0.118974
180704	332664	343	-0.06623	179414	331494	643	0.115253
181153	332925	218	-0.26741	179334	331366	801	0.343841
181147	332823	200	-0.29194	179255	331264	784	0.315048
181167	332778	194	-0.2738	179470	331125	1060	0.911576
181008	332777	207	-0.35854	179692	330933	119	-0.93882
180973	332687	180	-0.47198	179852	330801	778	1.177099
180916	332753	240	-0.28172	179140	330955	703	0.301245
181352	332946	180	-0.28474	179128	330867	676	0.308817
181133	332570	208	-0.09898	179065	330864	793	0.410891
180878	332489	198	-0.33577	179007	330727	685	0.300089
180829	332450	250	-0.12315	179110	330758	593	0.232742
180954	332399	192	-0.23533	179032	330645	549	0.15651
180956	332318	213	-0.07618	179095	330636	680	0.43598
180710	332330	321	0.09357	179058	330510	539	0.251791
180632	332445	569	0.516404	178810	330666	560	-0.0473

x	y	Zn	lzinc.res	x	y	Zn	lzinc.res
180530	332538	833	0.73979	178912	330779	1136	0.681844
180478	332578	906	0.748256	178981	330924	1383	0.848099
180383	332476	1454	1.198829	179076	331005	1161	0.709446
180494	332330	298	-0.1849	180151	330353	1672	2.520367
179211	331175	765	0.30767	180862	333116	415	-0.02467
181118	333214	279	-0.24449	180700	332882	474	0.109592
179474	331304	241	-0.684	180201	331160	126	-0.55046
179559	331423	317	-0.40811	180173	331923	210	-0.56964
179022	330873	545	-0.01072	180923	332874	220	-0.44197
178953	330742	505	-0.0657	180467	331694	133	-0.59743
178875	330516	420	-0.17457	179917	331325	141	-0.81526
178803	330349	332	-0.36752	179822	331242	158	-0.73642
179029	330394	400	0.002689	179991	331069	129	-0.66532
178605	330406	553	-0.08202	179120	330578	206	-0.69633
178701	330557	577	-0.04847	179034	330561	451	0.017205
179547	330245	155	-0.3575	179085	330433	296	-0.27123
179301	330179	224	-0.17819	179236	330046	189	-0.32175
179405	330567	180	-0.55466	179456	330072	154	-0.3358
179462	330766	226	-0.40455	179550	329940	169	-0.0669
179293	330797	186	-0.77951	179445	329807	403	0.79068
179180	330710	198	-0.76636	179337	329870	471	0.802963
179206	330398	187	-0.59304	179245	329714	612	1.080848
179618	330458	199	-0.18109	179024	329733	601	0.841323
179782	330540	157	-0.31729	178786	329822	783	0.822211
179980	330773	203	-0.02698	179135	329890	258	-0.00303
180067	331185	143	-0.56703	179030	330082	214	-0.41595
180162	331387	136	-0.66075	179184	330182	166	-0.5904
180451	331473	117	-0.59488	179085	330292	496	0.338037
180328	331158	113	-0.53801	178875	330311	342	-0.24472
180276	330963	130	-0.31832	179466	330381	162	-0.47963
180114	330803	192	0.024143	180627	330190	375	1.582918
179881	330912	240	-0.04484	179797	331919	139	-1.33496
179774	330921	221	-0.23437	179642	331955	253	-0.90628
179657	331150	140	-0.95259	179849	332142	703	0.187898
179731	331245	128	-1.03496	180265	332297	832	0.647208
179717	331441	166	-0.91759	180107	332101	262	-0.52824
179446	331422	191	-1.02086	180462	331947	142	-0.70365
179524	331565	232	-0.84706	180478	331822	119	-0.78274
179644	331730	203	-0.97608	180347	331700	152	-0.58126
180321	330366	722	1.83269				
180162	331837	210	-0.52328				
180029	331720	198	-0.63059				

Table D.4 **Dataset 2**

x x-coordinate
y y-coordinate
z simulated random field data with spherical covariance

x	y	z	x	y	z	x	y	z
1	1	7.17402	2	17	11.14071	4	12	9.765067
1	2	5.629593	2	18	13.07665	4	13	7.831999
1	3	8.062367	2	19	11.36171	4	14	9.305262
1	4	10.67703	2	20	8.34676	4	15	5.279547
1	5	9.743716	2	21	5.747576	4	16	9.045904
1	6	8.824026	3	1	5.774626	4	17	8.74689
1	7	8.452097	3	2	5.840706	4	18	8.656921
1	8	6.516037	3	3	7.474069	4	19	5.884602
1	9	5.090688	3	4	9.067919	4	20	9.343737
1	10	8.840694	3	5	9.43302	4	21	6.069462
1	11	7.403631	3	6	11.20671	5	1	8.848894
1	12	10.43596	3	7	15.21776	5	2	5.421146
1	13	8.430479	3	8	10.37691	5	3	7.630754
1	14	4.037384	3	9	11.98001	5	4	8.972671
1	15	9.072597	3	10	9.698153	5	5	12.65216
1	16	7.806457	3	11	11.73494	5	6	15.66395
1	17	10.96826	3	12	11.48687	5	7	16.62842
1	18	14.14552	3	13	4.389889	5	8	19.14759
1	19	10.80014	3	14	6.79473	5	9	14.54305
1	20	8.403758	3	15	6.974289	5	10	11.10615
1	21	7.696351	3	16	7.824554	5	11	8.638796
2	1	4.052948	3	17	12.56725	5	12	6.93789
2	2	2.677498	3	18	10.78082	5	13	6.370481
2	3	5.831359	3	19	8.915528	5	14	11.25082
2	4	7.829564	3	20	8.091145	5	15	7.391582
2	5	8.165773	3	21	4.445028	5	16	12.26019
2	6	10.67653	4	1	9.801252	5	17	9.064508
2	7	8.865751	4	2	6.833279	5	18	10.41991
2	8	10.06918	4	3	6.770902	5	19	8.78039
2	9	9.174273	4	4	7.733577	5	20	6.10727
2	10	7.08664	4	5	11.14517	5	21	1.150962
2	11	11.04742	4	6	16.16089	6	1	5.324075
2	12	11.29363	4	7	15.14026	6	2	9.02818
2	13	6.581363	4	8	17.73221	6	3	7.589614
2	14	5.877923	4	9	12.40462	6	4	9.754815
2	15	5.364476	4	10	11.29019	6	5	12.4134
2	16	7.903336	4	11	10.07988	6	6	14.77951

x	y	z	x	y	z	x	y	z
6	7	13.03711	8	7	13.94484	10	7	9.02584
6	8	15.21801	8	8	11.22395	10	8	11.65709
6	9	13.01373	8	9	4.486485	10	9	10.26607
6	10	7.255538	8	10	3.955301	10	10	10.17594
6	11	8.115974	8	11	7.990984	10	11	6.962769
6	12	7.793379	8	12	8.082298	10	12	3.618402
6	13	10.46487	8	13	7.520473	10	13	4.157749
6	14	9.631969	8	14	10.42724	10	14	9.057605
6	15	13.30261	8	15	12.0094	10	15	9.684853
6	16	12.17639	8	16	10.61203	10	16	10.7936
6	17	9.137557	8	17	14.1979	10	17	15.90126
6	18	7.930402	8	18	13.98177	10	18	10.85851
6	19	4.697162	8	19	12.1481	10	19	12.27293
6	20	3.160791	8	20	13.58151	10	20	16.35769
6	21	1.114256	8	21	9.354367	10	21	9.643144
7	1	9.61834	9	1	5.762639	11	1	9.139911
7	2	8.705332	9	2	4.597707	11	2	8.75564
7	3	10.51641	9	3	4.715583	11	3	7.98993
7	4	12.69328	9	4	11.44568	11	4	8.397481
7	5	12.21195	9	5	14.67201	11	5	9.620677
7	6	13.50481	9	6	13.46622	11	6	11.34102
7	7	12.16325	9	7	12.59351	11	7	7.96939
7	8	10.4961	9	8	11.30479	11	8	9.139277
7	9	6.741525	9	9	9.713824	11	9	8.201376
7	10	7.47939	9	10	6.120818	11	10	9.384007
7	11	7.291897	9	11	3.176536	11	11	7.829228
7	12	6.267766	9	12	4.360367	11	12	7.253819
7	13	10.19206	9	13	4.886309	11	13	6.563544
7	14	11.329	9	14	11.38052	11	14	10.05077
7	15	13.31622	9	15	8.841195	11	15	8.381211
7	16	12.36429	9	16	6.792326	11	16	11.01255
7	17	11.88702	9	17	13.15408	11	17	12.53181
7	18	12.35165	9	18	14.62528	11	18	9.714665
7	19	8.317193	9	19	13.35205	11	19	11.91186
7	20	5.669758	9	20	12.39708	11	20	11.56933
7	21	7.234667	9	21	10.94478	11	21	8.608255
8	1	6.034727	10	1	4.062228	12	1	10.68383
8	2	8.543246	10	2	3.223105	12	2	6.387075
8	3	9.507178	10	3	6.85874	12	3	8.571195
8	4	10.87592	10	4	10.48115	12	4	9.798716
8	5	16.33758	10	5	11.3921	12	5	10.89625
8	6	13.39103	10	6	11.27145	12	6	12.06791

x	y	z	x	y	z	x	y	z
12	7	9.059298	14	7	9.889018	16	7	4.057996
12	8	12.81476	14	8	7.415837	16	8	2.338483
12	9	7.378364	14	9	7.002544	16	9	2.105284
12	10	9.61742	14	10	5.317559	16	10	2.617382
12	11	9.362107	14	11	9.213404	16	11	4.27385
12	12	8.948381	14	12	10.3193	16	12	9.734679
12	13	12.27096	14	13	8.990375	16	13	6.760119
12	14	6.736837	14	14	9.017094	16	14	10.64087
12	15	7.03301	14	15	10.16623	16	15	7.646127
12	16	10.11089	14	16	10.70473	16	16	10.8493
12	17	12.03208	14	17	8.587085	16	17	9.263414
12	18	12.28911	14	18	7.456117	16	18	13.32071
12	19	9.394038	14	19	8.383049	16	19	9.627309
12	20	7.659285	14	20	10.15922	16	20	10.1726
12	21	8.877755	14	21	5.572544	16	21	8.248042
13	1	10.97172	15	1	11.26735	17	1	12.96496
13	2	9.963009	15	2	10.1774	17	2	10.43298
13	3	11.39525	15	3	9.991507	17	3	10.35202
13	4	8.583893	15	4	9.085094	17	4	4.025275
13	5	8.230277	15	5	7.374033	17	5	6.989688
13	6	9.721354	15	6	7.464331	17	6	5.520771
13	7	11.85987	15	7	8.114084	17	7	3.882676
13	8	10.9051	15	8	7.828652	17	8	3.571775
13	9	7.926802	15	9	6.701107	17	9	5.005649
13	10	7.331425	15	10	5.420708	17	10	4.785816
13	11	6.153275	15	11	8.725587	17	11	5.848873
13	12	13.25869	15	12	9.476228	17	12	9.456943
13	13	10.60707	15	13	10.65469	17	13	11.53219
13	14	10.12238	15	14	9.641443	17	14	7.258486
13	15	10.38767	15	15	9.758587	17	15	9.06499
13	16	8.722463	15	16	13.04432	17	16	10.39041
13	17	11.23824	15	17	11.77297	17	17	10.45395
13	18	10.85299	15	18	11.20498	17	18	8.92548
13	19	5.800037	15	19	8.923388	17	19	12.3095
13	20	9.196993	15	20	7.981092	17	20	11.57196
13	21	4.098469	15	21	6.800709	17	21	8.932247
14	1	13.40074	16	1	10.12634	18	1	13.01383
14	2	10.40635	16	2	9.878351	18	2	13.83675
14	3	13.02574	16	3	9.707438	18	3	10.77114
14	4	13.51192	16	4	9.503284	18	4	11.34262
14	5	10.55154	16	5	7.817091	18	5	9.789845
14	6	8.377151	16	6	7.78839	18	6	6.413784

x	y	z	x	y	z
18	7	8.580931	20	7	10.10014
18	8	8.420348	20	8	8.269807
18	9	7.181423	20	9	9.953446
18	10	7.280105	20	10	6.126669
18	11	8.145766	20	11	8.503288
18	12	9.673925	20	12	13.8138
18	13	10.80255	20	13	13.81252
18	14	12.26603	20	14	12.06784
18	15	13.8342	20	15	10.89617
18	16	11.97227	20	16	12.09027
18	17	13.01099	20	17	15.1767
18	18	14.54081	20	18	16.26678
18	19	14.24942	20	19	15.23952
18	20	12.63275	20	20	11.74135
18	21	12.95206	20	21	10.47022
19	1	11.82996	21	1	10.30439
19	2	12.56468	21	2	12.98922
19	3	9.560389	21	3	12.20233
19	4	10.05232	21	4	14.43337
19	5	9.609926	21	5	11.1502
19	6	9.583843	21	6	8.087771
19	7	13.27845	21	7	12.45687
19	8	9.562723	21	8	11.92957
19	9	8.656834	21	9	10.23353
19	10	5.439155	21	10	6.834537
19	11	7.029822	21	11	7.755994
19	12	9.9391	21	12	9.191904
19	13	11.01701	21	13	12.46046
19	14	11.05188	21	14	14.2694
19	15	9.861147	21	15	10.80741
19	16	14.43699	21	16	14.06963
19	17	13.69388	21	17	16.57438
19	18	14.51492	21	18	13.89543
19	19	17.21243	21	19	13.32699
19	20	14.12771	21	20	13.53778
19	21	10.96515	21	21	12.36474
20	1	10.59469			
20	2	11.25239			
20	3	10.72673			
20	4	12.30146			
20	5	11.35743			
20	6	12.13873			

Table D.5 Dataset 3

x x-coordinate
y y-coordinate
z simulated random field data with spherical covariance and superimposed trend

x	y	z	x	y	z	x	y	z
1	1	14.02039	2	17	91.28708	4	13	72.57837
1	2	16.97596	2	18	97.72302	4	14	78.55163
1	3	23.90873	2	19	100.5081	4	15	79.02591
1	4	31.0234	2	20	101.9931	4	16	87.29227
1	5	34.59008	3	1	15.22099	4	17	91.49326
1	6	38.17039	3	2	19.78707	4	18	95.90329
1	7	42.29846	3	3	25.92044	4	19	97.63097
1	8	44.8624	3	4	32.01429	4	20	105.5901
1	9	47.93706	3	5	36.87939	4	21	106.8158
1	10	56.18706	3	6	43.15308	5	1	20.89526
1	11	59.25	3	7	51.66412	5	2	21.96751
1	12	66.78233	3	8	51.32327	5	3	28.67712
1	13	69.27685	3	9	57.42637	5	4	34.51904
1	14	69.38375	3	10	59.64452	5	5	42.69853
1	15	78.91896	3	11	66.1813	5	6	50.21031
1	16	82.15282	3	12	70.43323	5	7	55.67478
1	17	89.81462	3	13	67.83626	5	8	62.69396
1	18	97.49189	3	14	74.7411	5	9	62.58942
1	19	98.64651	3	15	79.42066	5	10	63.65252
1	20	100.7501	3	16	84.77092	5	11	65.68516
1	21	104.5427	3	17	94.01362	5	12	68.48426
2	1	12.19931	3	18	96.72719	5	13	72.41685
2	2	15.32386	3	19	99.36189	5	14	81.79719
2	3	22.97773	3	20	103.0375	5	15	82.43795
2	4	29.47593	3	21	103.8914	5	16	91.80655
2	5	34.31214	4	1	20.54762	5	17	93.11088
2	6	41.3229	4	2	22.07965	5	18	98.96628
2	7	44.01212	4	3	26.51727	5	19	101.8268
2	8	49.71555	4	4	31.97994	5	20	103.6536
2	9	53.32064	4	5	39.89154	5	21	103.1973
2	10	55.73301	4	6	49.40726	6	1	18.67044
2	11	64.19378	4	7	52.88663	6	2	26.87455
2	12	68.94	4	8	59.97857	6	3	29.93598
2	13	68.72773	4	9	59.15098	6	4	36.60118
2	14	72.52429	4	10	62.53656	6	5	43.75977
2	15	76.51084	4	11	65.82625	6	6	50.62588
2	16	83.5497	4	12	70.01143	6	7	53.38348

x	y	z	x	y	z	x	y	z
6	8	60.06437	8	8	58.67032	10	8	61.70345
6	9	62.3601	8	9	56.43285	10	9	64.81244
6	10	61.10191	8	10	60.40167	10	10	69.22231
6	11	66.46234	8	11	68.93735	10	11	70.50914
6	12	70.63975	8	12	73.52867	10	12	71.66477
6	13	77.81123	8	13	77.46684	10	13	76.70412
6	14	81.47834	8	14	84.87361	10	14	86.10397
6	15	89.64898	8	15	90.95577	10	15	91.23122
6	16	93.02276	8	16	94.0584	10	16	96.83996
6	17	94.48392	8	17	102.1443	10	17	106.4476
6	18	97.77677	8	18	106.4281	10	18	105.9049
6	19	99.04353	8	19	109.0945	10	19	111.8193
6	20	102.0072	8	20	115.0279	10	20	120.4041
6	21	104.4606	8	21	115.3007	10	21	118.1895
7	1	24.26471	9	1	23.00901	11	1	28.98628
7	2	27.8517	9	2	26.34407	11	2	33.10201
7	3	34.16278	9	3	30.96195	11	3	36.8363
7	4	40.83964	9	4	42.19205	11	4	41.74385
7	5	44.85831	9	5	49.91837	11	5	47.46704
7	6	50.65118	9	6	53.21259	11	6	53.68738
7	7	53.80961	9	7	56.83988	11	7	54.81576
7	8	56.64247	9	8	60.05115	11	8	60.48564
7	9	57.38789	9	9	62.96019	11	9	64.04774
7	10	62.62576	9	10	63.86718	11	10	69.73037
7	11	66.93826	9	11	65.4229	11	11	72.6756
7	12	70.41413	9	12	71.10673	11	12	76.60019
7	13	78.83842	9	13	76.13268	11	13	80.40991
7	14	84.47536	9	14	87.12689	11	14	88.39714
7	15	90.96259	9	15	89.08756	11	15	91.22758
7	16	94.51066	9	16	91.53869	11	16	98.35891
7	17	98.53339	9	17	102.4004	11	17	104.3782
7	18	103.498	9	18	108.3716	11	18	106.061
7	19	103.9636	9	19	111.5984	11	19	112.7582
7	20	105.8161	9	20	115.1434	11	20	116.9157
7	21	111.881	9	21	118.1911	11	21	118.4546
8	1	21.98109	10	1	22.60859	12	1	31.8302
8	2	28.98961	10	2	26.26947	12	2	32.03344
8	3	34.45355	10	3	34.40511	12	3	38.71756
8	4	40.32229	10	4	42.52752	12	4	44.44508
8	5	50.28394	10	5	47.93847	12	5	50.04262
8	6	51.8374	10	6	52.31781	12	6	55.71428
8	7	56.89121	10	7	54.57221	12	7	57.20566

x	y	z	x	y	z	x	y	z
12	8	65.46113	14	8	62.6622	16	8	60.18485
12	9	64.52473	14	9	66.74891	16	9	64.45165
12	10	71.26379	14	10	69.56393	16	10	69.46375
12	11	75.50847	14	11	77.95977	16	11	75.62022
12	12	79.59475	14	12	83.56567	16	12	85.58105
12	13	87.41733	14	13	86.73674	16	13	87.10649
12	14	86.3832	14	14	91.26346	16	14	95.48724
12	15	91.17938	14	15	96.9126	16	15	96.99249
12	16	98.75726	14	16	101.9511	16	16	104.6957
12	17	105.1784	14	17	104.3335	16	17	107.6098
12	18	109.9355	14	18	107.7025	16	18	116.1671
12	19	111.5404	14	19	113.1294	16	19	116.9737
12	20	114.3057	14	20	119.4056	16	20	122.019
12	21	120.0241	14	21	119.3189	16	21	124.5944
13	1	33.41809	15	1	36.31372	17	1	40.61133
13	2	36.90938	15	2	39.72377	17	2	42.57934
13	3	42.84161	15	3	44.03787	17	3	46.99838
13	4	44.53026	15	4	47.63146	17	4	45.17164
13	5	48.67664	15	5	50.4204	17	5	52.63606
13	6	54.66772	15	6	55.0107	17	6	55.66714
13	7	61.30624	15	7	60.16045	17	7	58.52904
13	8	64.85147	15	8	64.37502	17	8	62.71814
13	9	66.37317	15	9	67.74747	17	9	68.65202
13	10	70.27779	15	10	70.96707	17	10	72.93218
13	11	73.59964	15	11	78.77195	17	11	78.49524
13	12	85.20506	15	12	84.0226	17	12	86.60331
13	13	87.05343	15	13	89.70105	17	13	93.17855
13	14	91.06875	15	14	93.18781	17	14	93.40485
13	15	95.83403	15	15	97.80495	17	15	99.71136
13	16	98.66883	15	16	105.5907	17	16	105.5368
13	17	105.6846	15	17	108.8193	17	17	110.1003
13	18	109.7994	15	18	112.7513	17	18	113.0718
13	19	109.2464	15	19	114.9698	17	19	120.9559
13	20	117.1434	15	20	118.5275	17	20	124.7183
13	21	116.5448	15	21	121.8471	17	21	126.5786
14	1	37.1471	16	1	36.47271	18	1	41.9602
14	2	38.65271	16	2	40.72472	18	2	47.28312
14	3	45.77211	16	3	45.05381	18	3	48.71751
14	4	50.75828	16	4	49.34965	18	4	53.78899
14	5	52.2979	16	5	52.16346	18	5	56.73621
14	6	54.62352	16	6	56.63476	18	6	57.86015
14	7	60.63538	16	7	57.40436	18	7	64.5273

x	y	z	x	y	z
18	8	68.86672	20	8	71.31617
18	9	72.12779	20	9	77.49981
18	10	76.72647	20	10	78.17304
18	11	82.09213	20	11	85.04965
18	12	88.12029	20	12	94.86017
18	13	93.74891	20	13	99.35889
18	14	99.7124	20	14	102.1142
18	15	105.7806	20	15	105.4425
18	16	108.4186	20	16	111.1366
18	17	113.9574	20	17	118.7231
18	18	119.9872	20	18	124.3131
18	19	124.1958	20	19	127.7859
18	20	127.0791	20	20	128.7877
18	21	131.8984	20	21	132.0166
19	1	42.07633	21	1	43.15076
19	2	47.31105	21	2	50.33558
19	3	48.80676	21	3	54.0487
19	4	53.79869	21	4	60.77974
19	5	57.85629	21	5	61.99657
19	6	62.33021	21	6	63.43414
19	7	70.52481	21	7	72.30323
19	8	71.30909	21	8	76.27594
19	9	74.9032	21	9	79.0799
19	10	76.18552	21	10	80.1809
19	11	82.27619	21	11	85.60236
19	12	89.68547	21	12	91.53827
19	13	95.26337	21	13	99.30682
19	14	99.79825	21	14	105.6158
19	15	103.1075	21	15	106.6538
19	16	112.1834	21	16	114.416
19	17	115.9402	21	17	121.4207
19	18	121.2613	21	18	123.2418
19	19	128.4588	21	19	127.1734
19	20	129.8741	21	20	131.8841
19	21	131.2115	21	21	135.2111
20	1	42.14106			
20	2	47.29876			
20	3	51.2731			
20	4	57.34783			
20	5	60.9038			
20	6	66.18509			
20	7	68.64651			

Appendix E

Soil Classification

E.1 The Canadian System of Soil Classification (CSSC)

Soils are identified and classified within a hierarchical taxonomic system in much the same way as plants and/or animals. The concept of soil properties (chemical and physical parameters) reflecting the action and interaction of soil forming processes (Subsection 3.6.1) over time was introduced in the mid to late 1800s in both Russia and the United States. This concept allowed soils to be classified and related based upon their soil properties not just the environmental factors that influence the soil (e.g., vegetation).

The first soil survey was conducted in Canada in Ontario in 1914 and by 1949 some level of soil survey had been completed for most provinces and territories in Canada. Although soils were being classified during these surveys, a standardized Canadian system of classification was not in place. The first meeting of the National Soil Survey Committee of Canada (NCSS) in 1945 identified the need for such a system and at a similar meeting in 1955 the first Canadian taxonomic system of soil classification was presented. Revisions of that classification system have resulted in the current Canadian System of Soil Classification (CSSC), as defined by the 1998 Soil Classification Working Group (SCWG) (AAFC, 1998).

E.1.1 Soil Taxonomy

Taxonomy is defined as the science of classification and is used to group or categorize individuals with similar characteristics. Soil classification and taxonomy are related to both quantitative and qualitative differences in physical and/or chemical properties among soils. Taxa within the CSSC are hierarchical and in order of increasing specificity include: Order, Great Group, Subgroup, Family, and Series.

The Soil Order reflects both the effects of the dominant soil forming processes and/or environmental factors (Subsection 3.6.1). The Great Group reflects differentiation within an Order and is based upon differences in the strengths of the dominant soil processes and/or the effect(s) of a non-dominant soil process. Great Groups are further divided into Subgroups based upon the type and arrangement of soil horizons, and Subgroups can be divided into Families based upon parent material characteristics. The final and most specific taxon is the Soil Series.

The CSSC (AAFC, 1998) defines a Series as containing soils that "have similar kinds and arrangements of horizons, whose colour, texture, structure consistence, thickness, reaction and composition fall within a narrow range". Soils within a Series are identified by a specific name, which is typically linked to a geographic area; the name becomes representative of all of the characteristics of a particular soil. At the field level, a soil series is associated with the soil profile, its classification, and the location of that profile within the topographic landscape.

E.1.2 Soil Pedon

A soil pedon is generally understood to be the smallest, three-dimensional unit that can be considered a soil. The CSSC (AAFC, 1998) defines the lateral dimension of a pedon based upon genetic soil horizons. If soil horizons either vary or are faintly expressed and can be sampled within a lateral distance of 1 m, then that is the lateral extent of the pedon. If horizons are cyclical or intermittent and repeated within 2 to 7 m, then the lateral extent of the pedon is half the distance of the cycle (1 to 3.5 m). The vertical extent of the pedon is to the depth of the control section. The control section typically extends from the surface of mineral soils to either 25 cm below the upper boundary of the C, IIC, or permafrost table or to a depth of 2 m, whichever is less. Exceptions to this are further discussed in the CSSC (AAFC, 1998). Figures E.1 and E.2, reproduced by permission of the National Research Council of Canada (NRC) Research Press (AAFC, 1998), illustrate two examples of soil pedons.

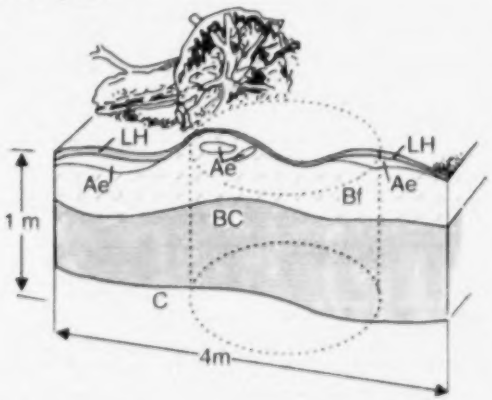


Figure E.1 Pedon of Podzolic soils within a hummocky terrain.

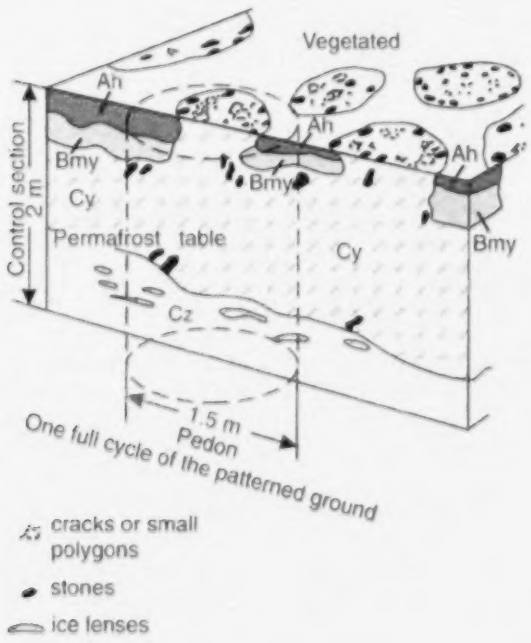


Figure E.2 Pedon of Cryosolic soils in an area of non-sorted circles.

E.1.3 Soil Horizons and Profile

As outlined in Subsections E.1.1 and E.1.2, soils are classified based upon the similarities and/or differences of their soil properties. These properties reflect the action and interaction of processes over time; the cumulative effect of these processes also results in the development or degradation of soil horizons. Soil horizons are typically observed within soil profiles as lateral layers of mineral or organic material. The primary mineral horizons are defined as A, B, and C; the primary organic horizons are L, F, and H (predominantly forest litter), or O (predominantly wetland vegetation). Soil horizons can be further defined by adding lower case or numerical suffixes to the primary horizon designation[58] (e.g., Ah). A full list of suffixes can be found in the CSSC (AAFC, 1998).

A soil profile is typically what is observed and/or sampled in the field to determine the classification of a soil. An example of a soil profile is included below as Figure E.3.
The CSSC (AAFC, 1998) also provides diagrammatic representations of typical soil profiles observed for each Soil Order.

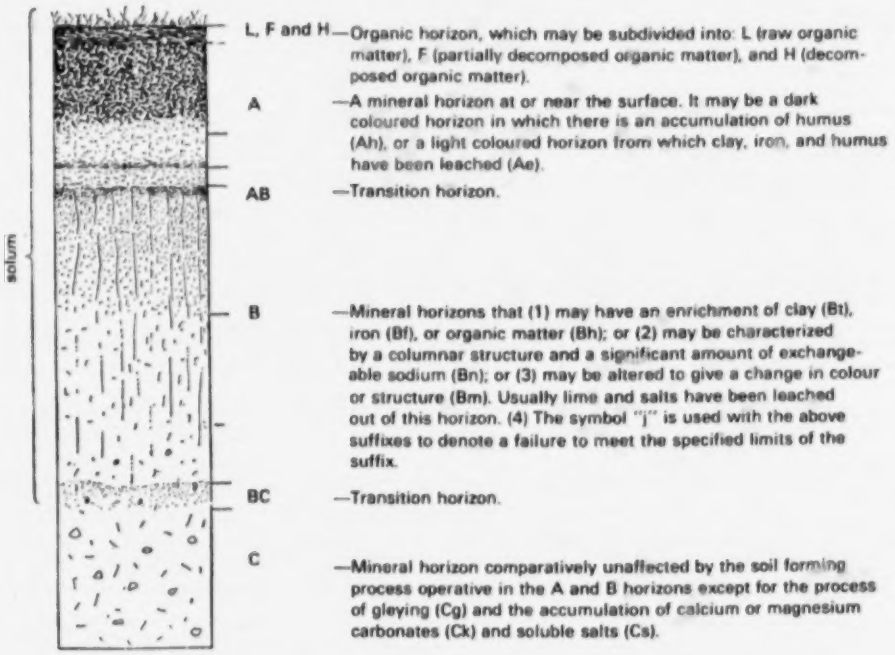


Figure E.3 Example of a soil profile (University of Alberta, 1975).

[58] Some regions in Canada may have more detailed levels of classification in addition to systems that build upon the CSSC. An example is British Columbia (see Luttmerding *et al.* 1990).

E.2 Canadian Soil Orders

The CSSC (AAFC, 1998) identifies and defines the following 10 Soil Orders within Canadian landscapes: Brunisolic, Chernozemic, Crysolic, Gleysolic, Luvisolic, Organic, Podzolic, Regosolic, Solonetzic, and Vertisolic. Most of these Orders can be said to predominantly occur in one of the following systems: forest, grassland, or tundra; therefore, they have a defined geographic extent. Table E.1 organizes the Soil Orders by the system in which they predominantly occur and summarizes their diagnostic horizon(s) and descriptions. Photographs of two soil profiles for each Soil Order can be found in Figure E.4. The geographical extent of each Soil Order is depicted in Figure E.5 which follows the same organization as Table E.1.

Table E.1 A summary of the Canadian System of Soil Classification Soil Orders, where they predominantly occur, their diagnostic horizons and descriptions.

Soil Order	Predominant occurrence	Diagnostic horizon(s)[1]	Description/Comments[2]
Brunisolic	Forest soils	B horizon (Bm)	Although more developed than Regosolic soils, Brunisolic soils cannot meet the requirements of any other Soil Order. These weakly developed soils are typically found on sandy materials with low clay and iron contents.
Luvisolic	Forest soils	B horizon (Bt)	The diagnostic Bt horizon develops from clay movement within the soil profile from the A horizon (eluviation) to the B horizon (illuviation). These soils tend to develop on parent materials that originated from sedimentary rock, which tends to be calcareous.
Podzolic	Forest soils	B horizon (Bf, Bh, Bhf)	The diagnostic Podzolic horizon develops from the deposition of iron and aluminum within the B horizon. Surficial soils tend to be acidic and predominantly develop under coniferous forest vegetation. These soils tend to be more weathered and leached than Luvisolic soils and develop on parent materials derived from igneous rocks.
Organic	Boreal forest/ Wetland soils	O horizon (Of, Om, Oh)	Organic soils develop from organic materials and tend to be situated in saturated to poorly drained conditions. Organic soils are at least 17% organic carbon by weight and have at least 40 to 60 cm of organic material at the surface or at least 10 cm if overlying a lithic contact (unconsolidated rock).
Crysolic	Subarctic forest / Arctic and tundra soils	B (By, Bz), C (Cy, Cz) and O (Oz) horizons	Cryosolic soils are identified as having permafrost within either 1 m of the surface or within 2 m of the surface if the soil profile shows evidence of cryoturbation (y), the mixing of the soil due to frost action. These soils can be organic or mineral.

Gleysolic	All regions	B (Bg) and C (Cg) horizons and in some cases the A (Aeg) horizon	Gleysolic soils can develop in any poorly drained and/or saturated mineral soil. Gleying and mottling are the diagnostic features and develop due to reducing conditions (anaerobic) within the soil profile. At least on of these features must occur within the upper 50 cm of soil in a band at least 10 cm thick.
Regosolic	All regions	No diagnostic horizon	Regosolic soils are identified as lacking the development of a B horizon to the extent that it either does not exist or is < 5-cm thick. These soils predominantly display characteristics that are similar to the parent material upon which they formed.
Chernozemic	Grassland and grassland-forest transition soils	A horizon (Ah, Ap)	The diagnostic Chernozemic horizon has an accumulation of organic matter from vegetative decomposition and must be at least 10-cm thick. Most Chernozemic soils are in agricultural production within western Canada.
Solonetzic	Grassland soils	B horizon (Bn, Bnt)	The diagnostic Solonetzic horizon has a ration of exchangeable Ca to Na of < 10. This horizon is typically prismatic or columnar and is very hard when dry and has low permeability due to swelling (high clay content) when wet. Solonetzic soils tend to either develop on saline parent material or have salts introduced by saline waters.
Vertisolic	Grassland soils	B (Bv, Bss) and C (Css) horizons	Vertisolic soils are highly localized within Canada, occurring in areas with high clay contents (≥ 60%). The effects of wet/dry, swelling/shrinking processes within the soil result in highly churned soils and or visible sloughing of surface material into cracks within the soil (vertic horizon, v). In addition to the vertic horizon, these soils must also contain slickensides within the upper 1 m of soil.

[1]Full descriptions of Soil Orders can be found within the Canadian System of Soil Classification (AAFC, 1998).
[2]This table was compiled from the following reference material: Agriculture Canada (1977). Agriculture and Agri-Food Canada (1998), and University of Saskatchewan (2009).

Most of the following photographs of soil profiles of the 10 Canadian soil Orders (Figure E.4) were reproduced courtesy of the Soil Landscapes of Canada Working Group (2007); the two photographs of Vertisolic soils have been reproduced with permission from the (NRC) Research Press (AAFC, 1998). Each soil order is presented with photographs of two different soil profiles. The geographical extent for each Soil Order, reproduced courtesy of the University of Saskatchewan (University of Saskatchewan, 2009), is presented in Figure E.5.

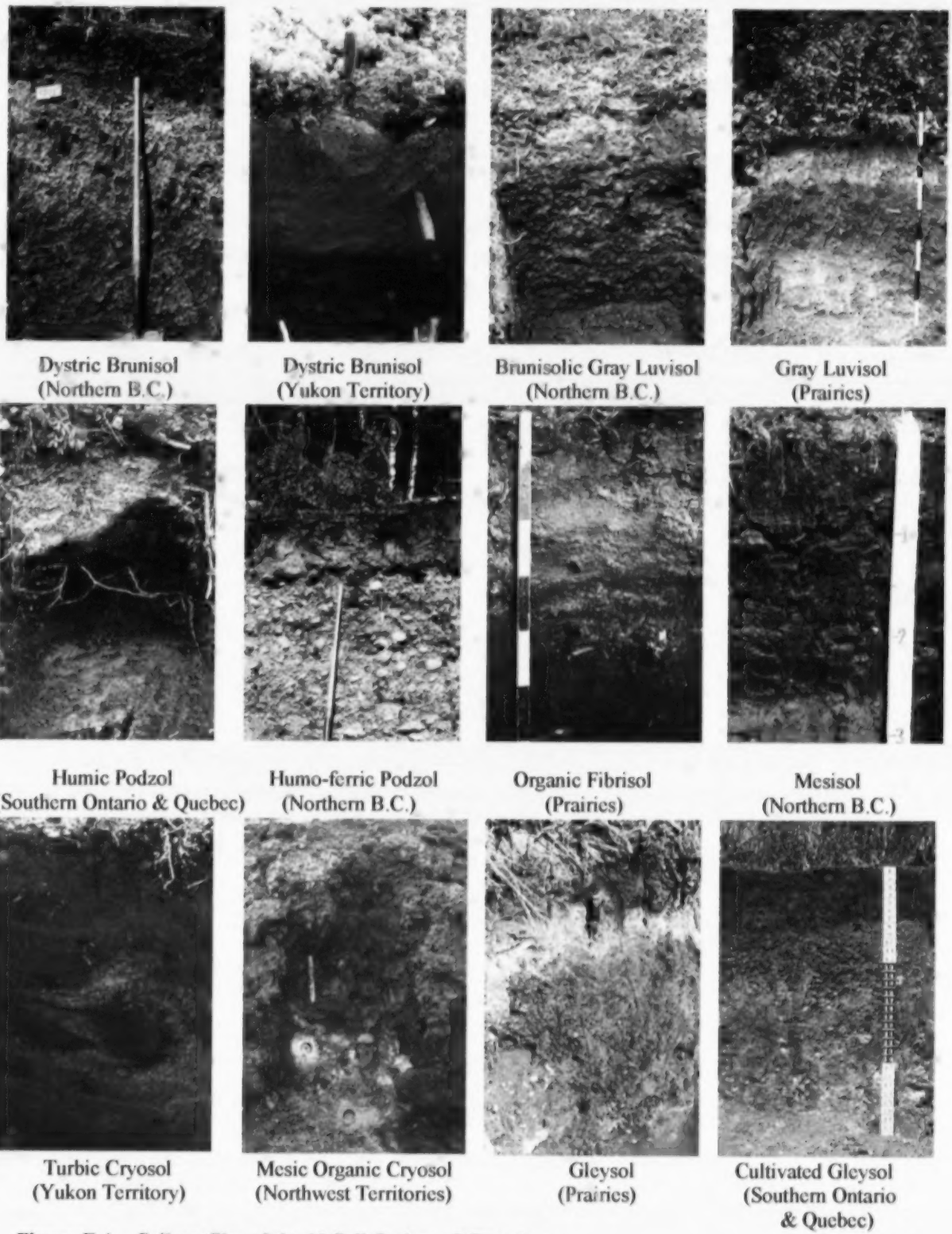
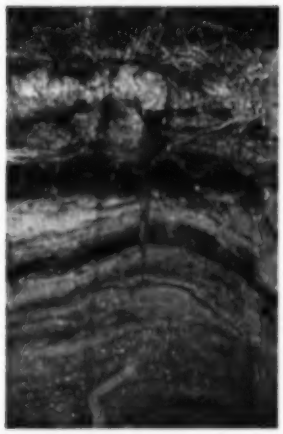


Figure E.4 Soil profiles of the 10 Soil Orders of Canada



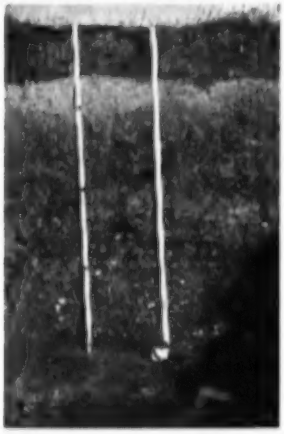
Cumulic Regosol
(Northern B.C.)

Regosol
(Atlantic Provinces)

Dark Brown Chernozem
(Prairies)

Black Chernozem
(Southern B.C)

Solodized Solonetz
(Prairies)



Black Solonetzic Profile
(Northern B.C.)

Orthic Humic Vertisol
(Manitoba)

Orthic Vertisol
(Saskatchewan)

Figure E.4 Soil profiles of the 10 soil Orders of Canada (con't).

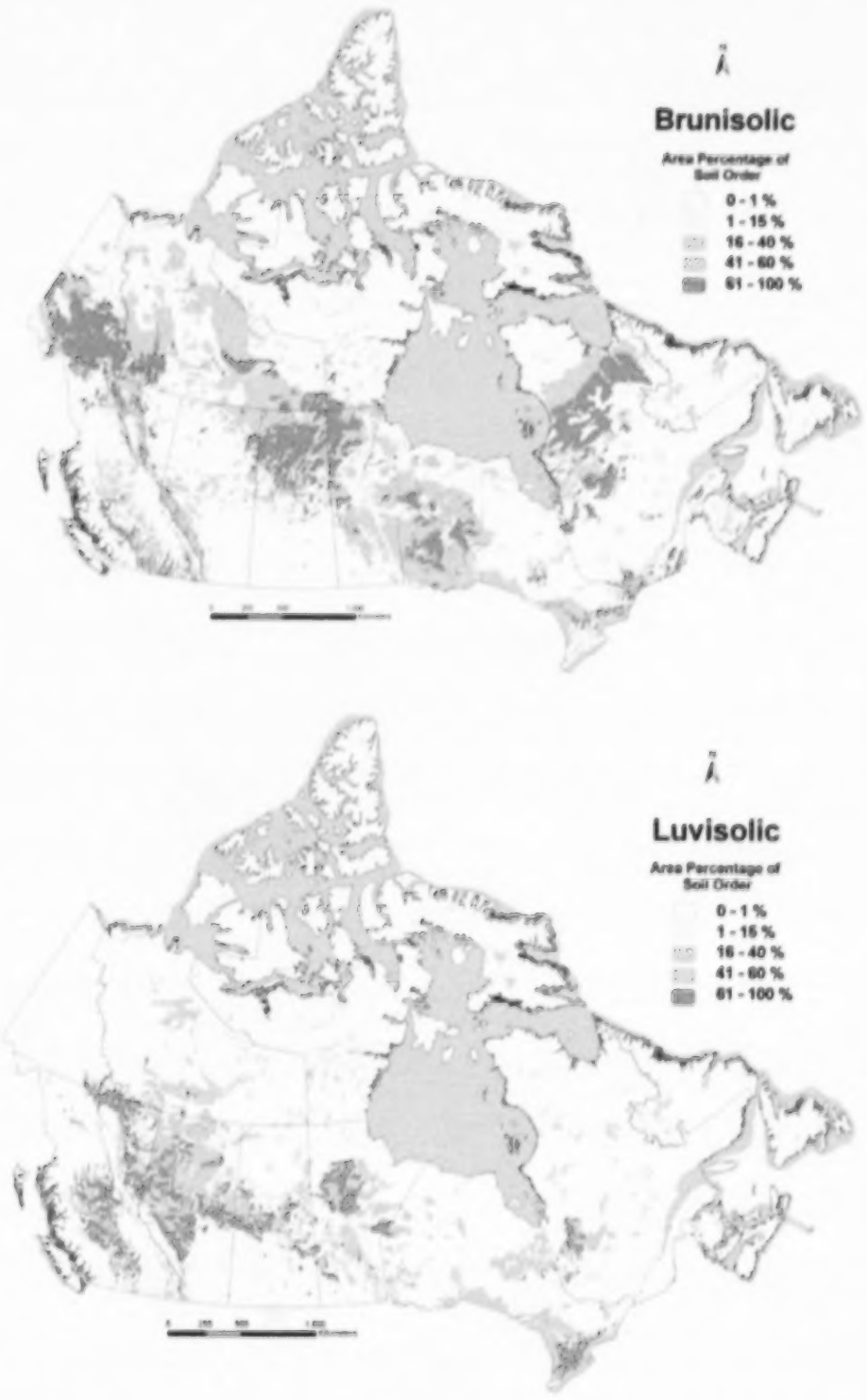
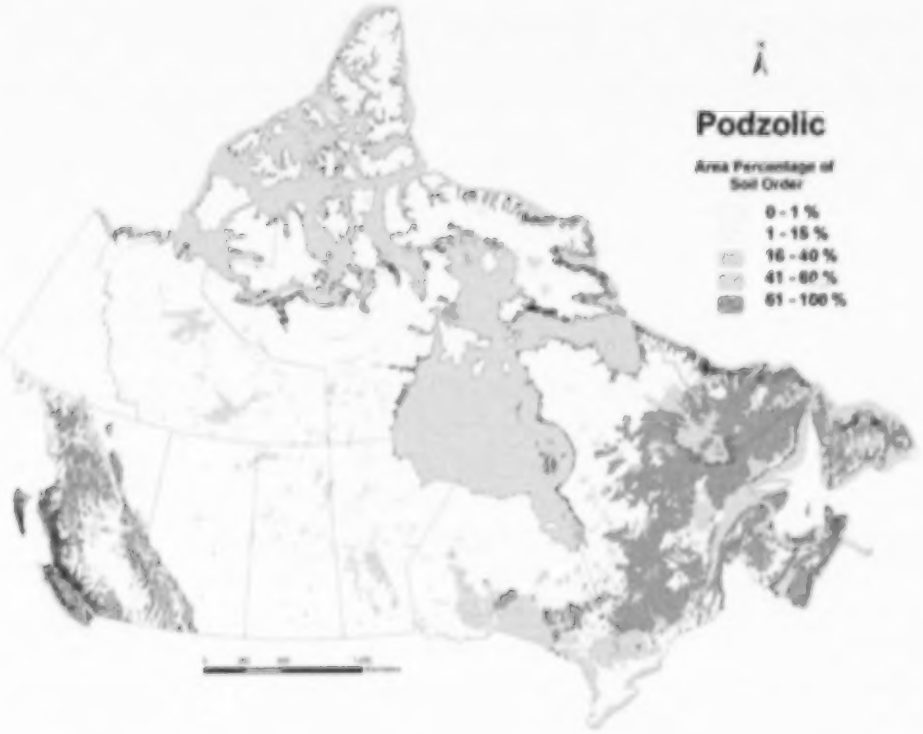


Figure E.5 Geographic extent of the 10 Canadian Soil Orders.



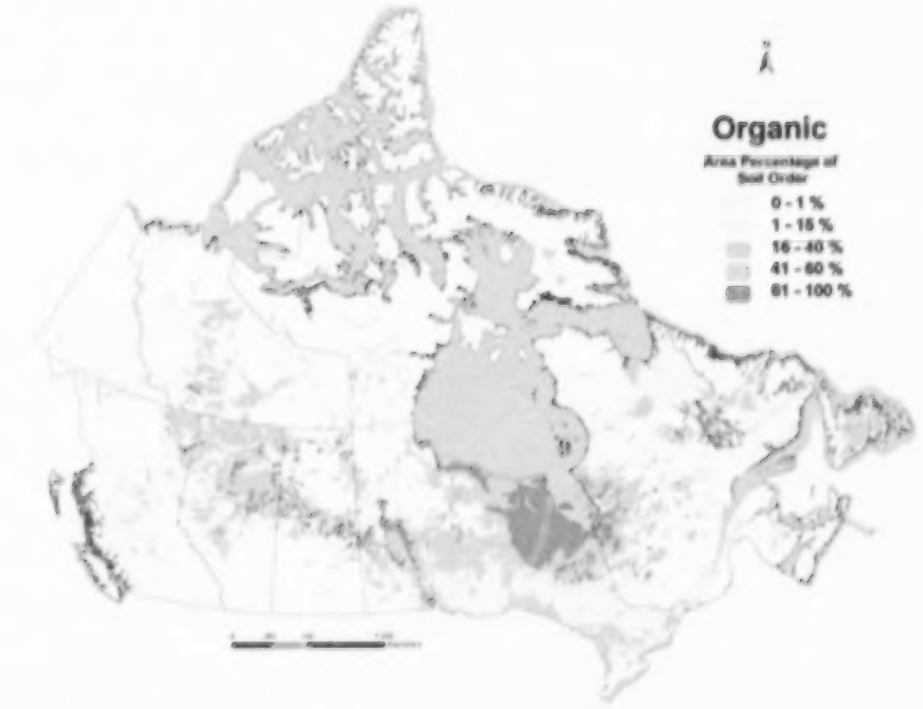


Figure E.5 Geographic extent of the 10 Canadian Soil Orders (con't).

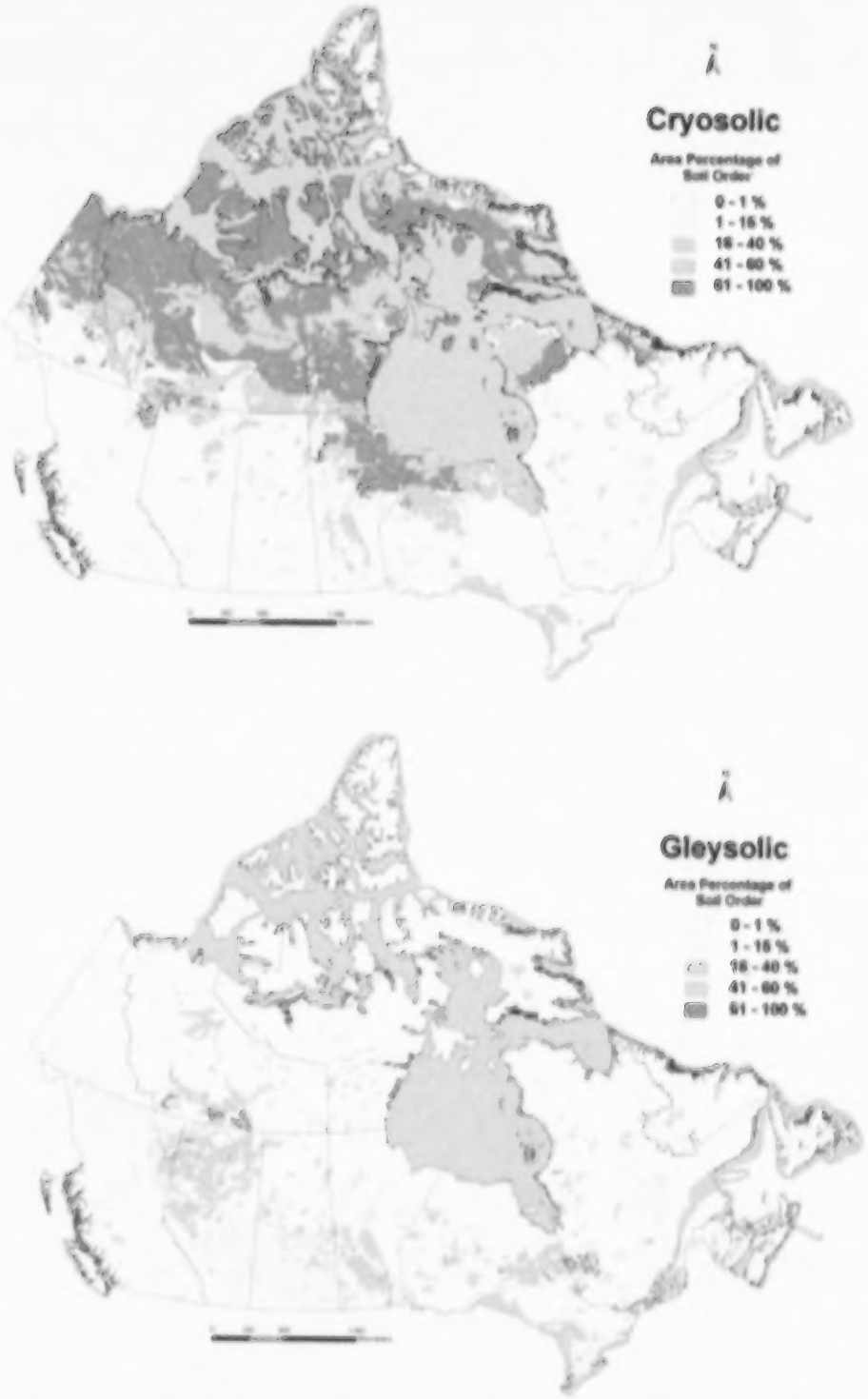
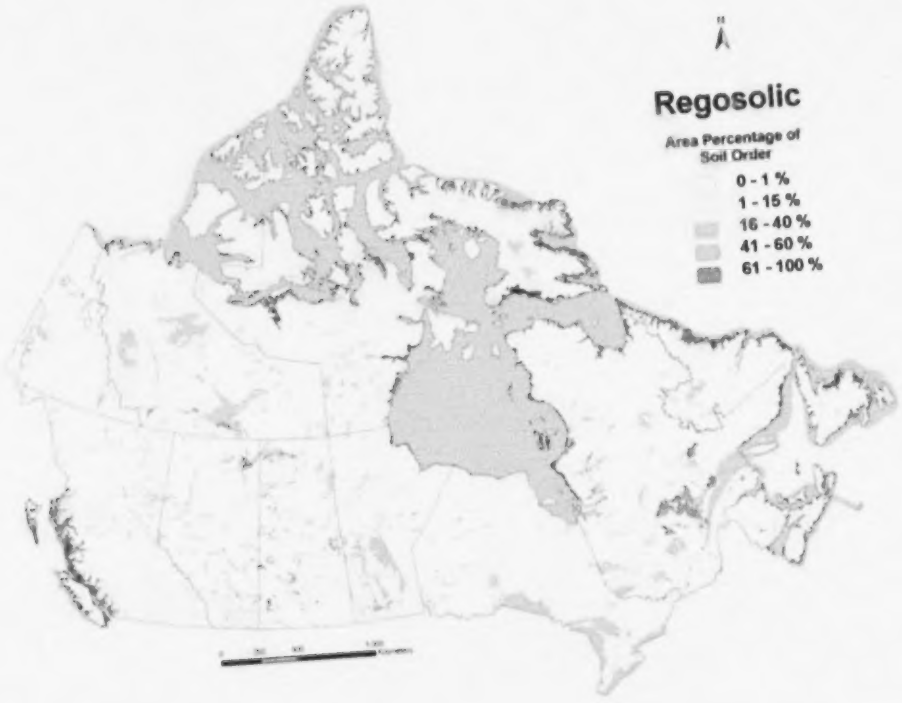


Figure E.5 Geographic extent of the 10 Canadian soil Orders (con't).



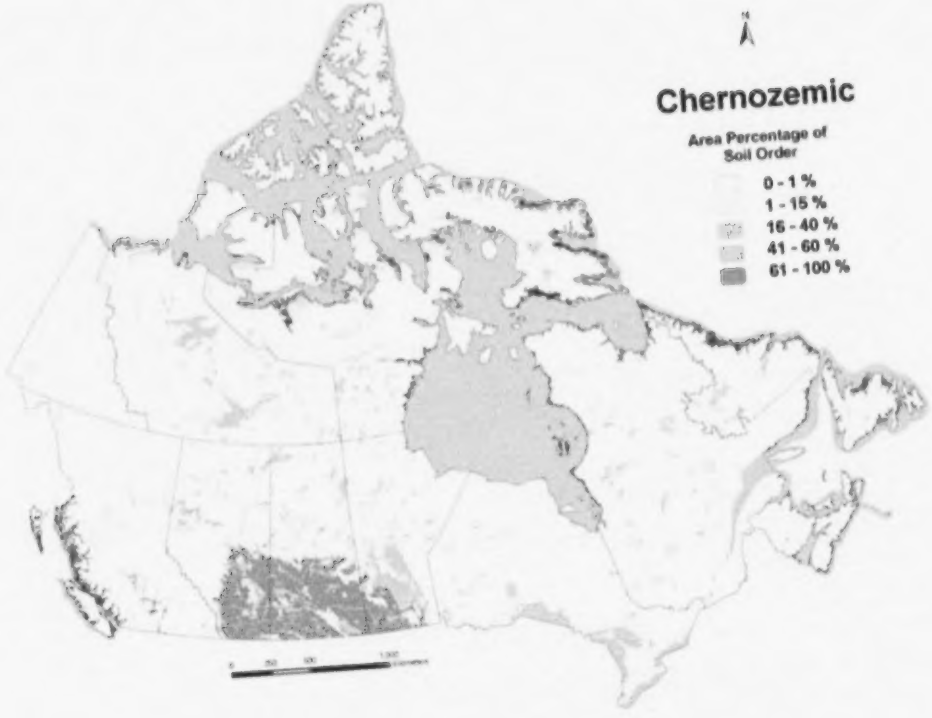


Figure E.5 Geographic extent of the 10 Canadian Soil Orders (con't).

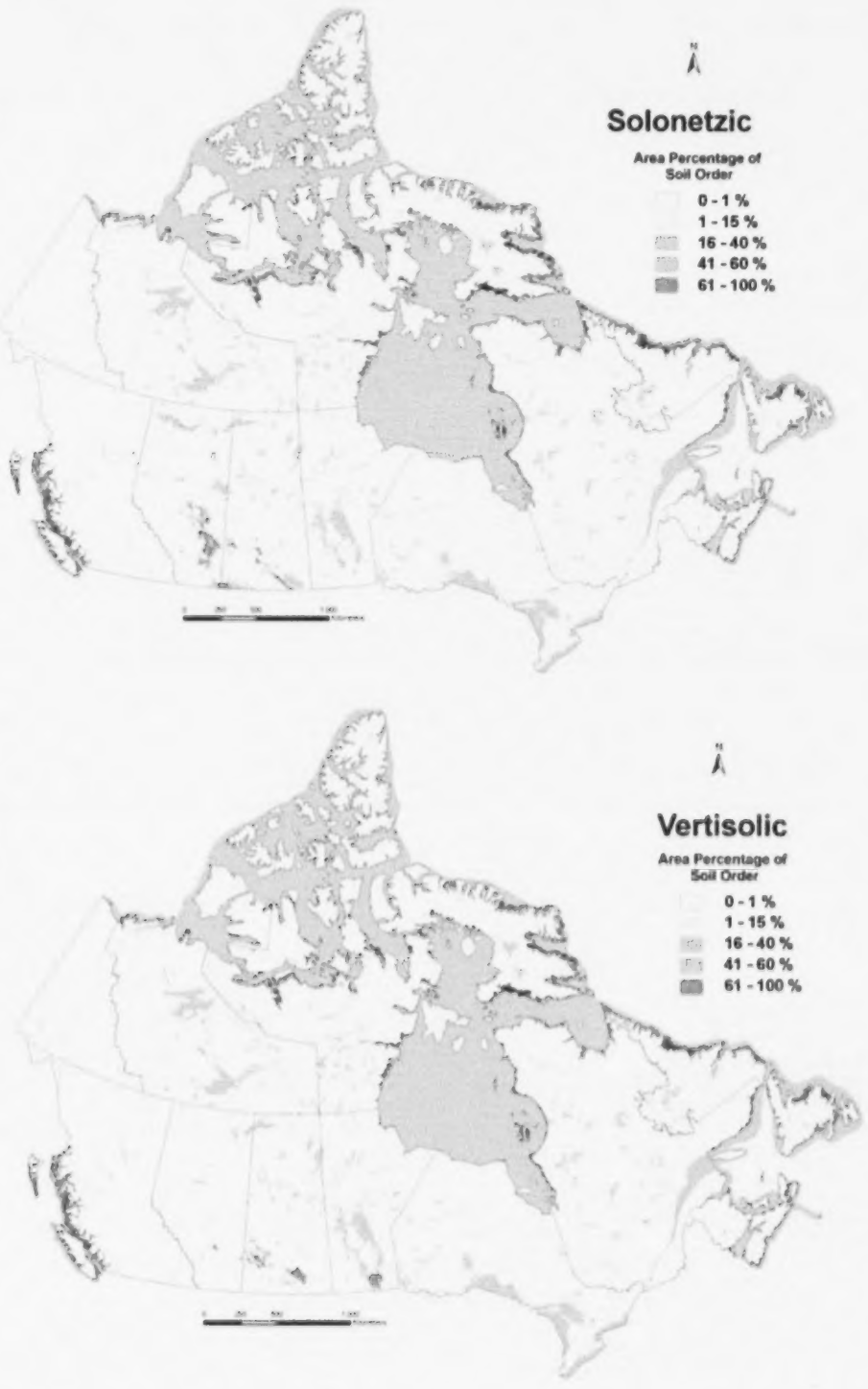


Figure E.5 Geographic extent of the 10 Canadian Soil Orders (con't).

Appendix F

Useful Web sites

Table F.1 A summary of web-based resources that can be used to obtain background soil, geological, geographical and ecological data in support of biological testing of soil from contaminated sites and other useful information.

Web site	Related section	Web site Address
Agri-Geomatics - Provides geographic information online to improve decision making and risk management for agriculture and the environment. (Agriculture Canada)	3.3	http://www4.agr.gc.ca/AAFC-AAC/display-afficher.do?id=1226330737632&lang=eng
Forest Ecosystem Mapping of Canada (Natural Resources Canada) This site also includes links to maps: *Interactive maps (From this site):* • Snapshots of disturbances • Net primary production (NPP) • Vegetation zones and bioclimatic domains of Quebec *From CFSNet:* • Forest ecosystem mapping in Canada • *From The Atlas of Canada:* • Wetlands • Wetland diversity • Terrestrial ecozones • Land cover diversity • Land cover • Productive forest land use	3.3; 6; B; D;	https://pfc.cfsnet.nfis.org/mapserver/ecoclimatic_pub/htdocs/ecoclimatic_e.phtml http://ecosys.cfl.scf.rncan.gc.ca/carte-map/carte-pert-map-dist-eng.asp http://ecosys.cfl.scf.rncan.gc.ca/dynamique-dynamic/productivite-productivity-eng.asp http://ecosys.cfl.scf.rncan.gc.ca/carte-map/bioclimatique-bioclimatic-eng.asp https://pfc.cfsnet.nfis.org/mapserver/ecoclimatic_pub/htdocs/ecoclimatic_e.phtml http://atlas.nrcan.gc.ca/site/english/maps/freshwater/distribution/wetlands http://atlas.nrcan.gc.ca/site/english/maps/environment/ecology/components/wetlanddiversity http://atlas.nrcan.gc.ca/site/english/maps/environment/ecology/framework/terrestrialecozones http://atlas.nrcan.gc.ca/site/english/maps/environment/ecology/components/landcoverdiversity http://atlas.nrcan.gc.ca/site/english/maps/environment/land/landcover http://atlas.nrcan.gc.ca/site/english/maps/environment/ecology/humanactivities/productiveforestlanduse
CanSIS (Canadian Soil Information System) (Agriculture and Agri-Food Canada) • Soil Landscapes of Canada • A National Ecological Framework for Canada • National Soil Database	3.3; 3.6; B; D	http://sis.agr.gc.ca/cansis/intro.html http://sis.agr.gc.ca/cansis/nsdb/slc/intro.html http://sis.agr.gc.ca/cansis/nsdb/ecostrat/intro.html http://sis.agr.gc.ca/cansis/nsdb/index.html

Web site	Related section	Web site Address
Geological Society of Canada - provides mapping tools – e.g., topographical maps (Natural Resources Canada) • General land maps • More specific terrestrial maps	3.3; D	http://ess.nrcan.gc.ca/index_e.php http://ess.nrcan.gc.ca/mapcar/index_e.php
Transportation of Dangerous Goods (Transport Canada)	3.8; 4.3; E	http://www.tc.gc.ca/eng/tdg/menu.htm
Terrestrial Ecozones of Canada (Parks Canada)	6; B	http://www.pc.gc.ca/apprendre-learn/prof/itm2-crp-trc/pdf/ecozone_e.pdf
Canada Centre for Remote Sensing – (Natural Resources Canada)	B	http://ccrs.nrcan.gc.ca/index_e.php
R Foundation for Statistical Computing	C	http://www.R-project.org

Appendix G

Soil Sampling Devices

A description and the advantages and disadvantages of the more commonly used soil collection devices available for use are presented in Table G.1. Diagrams and/or photographs of selected sampling devices are provided in Figure G.1 through Figure G.6. The information in Table G.1 is not exhaustive and further information can also be found in ASTM (2005, 2006b).

Table G.1 Advantages and disadvantages of different soil collection devices.

Device	Type of samples collected	Soil type	Soil area or volume sampled	Soil depth sampled	Advantages	Disadvantages	Reference(s)
Shovel, scoop, spoon, trowel, spade	Unconsolidated	All soil types including non-cohesive sandy or loose soils	~ 0.5 to 4 L	Surface, shallow subsurface	Easily and quickly collects large volumes of soil; collects blocks of soil; easy to decontaminate	Samples can be biased because of shape and imprecise volume but bias is minimized by careful sample collection	Mason, 1992; CCME, 1993a; Prévost and Antoun, 2008
Cutting/sampling frame	Unconsolidated	Organic horizon(s), mineral A horizon(s)	100 to 900 cm^2	Surface	Efficient way to collect representative bulk sample	May be difficult to remove all soil within frame	Bélanger and Van Rees, 2008
Reference frame + shears and trowel/scoop	Unconsolidated	Organic horizon(s), mineral A horizon(s)	100 to 900 cm^2	Surface	Efficient way to collect representative bulk sample	May be difficult to remove all soil within frame	Bélanger and Van Rees, 2008
Ring sampler	Consolidated or unconsolidated	All cohesive soils	0.5 to 30 cm diameter	Surface 0 to 20 cm	Easy to use, precise core	Not as useful for unconsolidated soils or hard clay	ISO, 2002b; Mason, 1992
Bulb planter	Consolidated or unconsolidated	All cohesive soils	7.5 to 5.5 × 11.5 cm ~ 1.5 L	0 to 15 cm	Large core, better if large volume needed	Not useful for hard soils	Dalpé and Hamel, 2008; CCME, 1993a
Cutting cylinder (soil punch)	Consolidated or unconsolidated	Organic, A horizon	59 to 556 cm^2	Surface	Soil cores are large and can efficiently sample large volumes	May slightly compress soil samples	Bélanger and Van Rees, 2008; ISO, 2003a

Soil coring device (manual)	Consolidated or unconsolidated	All cohesive soils	2.5 to 10 cm diameter; 30 to 60 cm long	0 to 60 cm	Easy to use, precise core, easily cleaned, can use liner or sample tube	Compaction when driving corer into soil. Cores not truly undisturbed unless liner used	CCME, 1993a; USEPA, 2006; Mason, 1992
Slide-hammer core sampler	Consolidated or unconsolidated	All cohesive soils	2.5 to 10 cm diameter; 30 to 60 cm long	0 to 60 cm	Easy to use, precise core, easily cleaned, can use liner or sample tube	Compaction when driving corer into soil. Cores not truly undisturbed unless liner used	EC and SRC, 2007
Auger (manual)	Unconsolidated	All cohesive soils	2.5 to 15 cm long	0 to 60 cm	Easy to use; modified versions for different soil types	Less precise sample than coring device; hard to decontaminate, modifies soil matrix, can introduce artefacts into soil sample	Mason, 1992
Split-spoon or tube sampler	Consolidated or unconsolidated	All cohesive soils including hard soils	Varied; up to 10 cm diameter; up to 2 kg per sample	0 to 40 cm or 0 cm to bedrock	Easy to use, precise core, can use liner, large cores	Deep cores can only be obtained using drilling rig	Weinfurtner and Kördel, 2007; CCME, 1993a; ASTM 2008a; ASTM 2009b
Shelby-tube sampler	Consolidated or unconsolidated	All cohesive soils including hard soils	Varied; up to 10 cm diameter	0 to 40 cm or 0 cm to bedrock	Easy to use, precise core, can use liner	Deep cores can only be obtained using drilling rig, not durable in hard soils	CCME, 1993a
Piston samplers	Consolidated or unconsolidated	Non-cohesive soils, wet soils, wet clay, dry and wet peat	Varied	Surface, shallow subsurface	Holds moisture and fine materials in place in sample	Somewhat difficult to operate	Mason, 1992; Sheppard *et al.*, 1993
Direct push corer tubes (agricultural or GeoProbe™)	Consolidated	All cohesive soils	Tubes 5 or 7 cm diameter and 1.2 m long;[1] size of probes and liners[2] varied	Surface	Saturated sands and silts can be collected, consolidated samples used to classify soils	Must use a drill rig, not optimal in wet conditions with stony soils or soils with a high clay content	Similar method to ASTM 2008a

Rotary (hollow stem auger with lined[2] or unlined core barrels)	Consolidated	All cohesive soils and soft bedrock	Varied	Surface to bedrock	Saturated sands and silts can be collected	Must use a drill rig, not suitable for stony soils, modifies soil matrix, can introduce artefacts into soil sample	ASTM 2008a; 2008b; 2008c; 2009b
Rotary (solid stem auger)	Unconsolidated	All cohesive soils, frozen soils, soft bedrock	Varied; for contaminant sampling diameter usually 15 cm but can be larger	Surface to bedrock	Easy to use, faster than hollow stem, provides continuous lithology information	Must use a drill rig, limited by stony soils, sample depth determination can be imprecise due to auger sample spin-up, modifies soil matrix, can introduce artefacts into soil sample	ASTM 2009a
Peat profile and core cutters	Consolidated or unconsolidated	Organic soil	Varied	Surface	Easy to use, cuts away core	Surface soil use only	Sheppard *et al.*, 1993
Chain saw	Consolidated or unconsolidated	Frozen organic soil	Varied	Surface, shallow subsurface	Easy to use with training	Potential for cross contamination from chainsaw oil	Sheppard *et al.*, 1993; Tarnocai, 1993
Pickaxe	Unconsolidated	Frozen soil, permafrost	Varied	Surface, subsurface	Easy to use	Potentially biased sample	USEPA, 2006; Tarnocai, 1993
Hoffer probe	Consolidated or unconsolidated	Frozen soil, frozen peat	2.6 to 4.3 cm diameter, 10 to 15 cm long cores	1 to 5 m	Light, portable, durable, one of few tools to sample frozen organic soils below 0.5 m	Sampling below 5 m difficult	Sheppard *et al.*, 1993; Tarnocai, 1993
Modified CRREL auger	Unconsolidated	Frozen soil, permafrost	3.8 to 7.5 diameter	1 to 5 m	Robust, easy to use	Core may remain in soil and needs to be extracted using a core catcher	Tarnocai, 1993

[1] Drilling equipment specifications originally provided in imperial units but have been converted to S.I. units in this table
[2] Liners can be Teflon, PVC, brass or stainless steel and vary in length from 2.54 cm to 1.2 m.

The following figures are photographs and/or diagrams of selected sampling devices described in Table G.1.

Figure G.1 Slide-hammer core sampler (A), sample location (B) and soil core once plastic sleeve is removed from the sampler (EC and SRC, 2007).

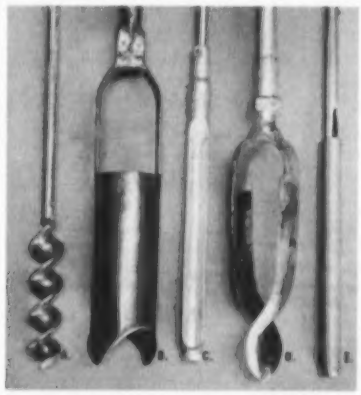


Figure G.2 Select examples of manual field sampling equipment including augers and probes (A. screw or worm; B. barrel; C. probe, D. "Dutch"; and E. peat samplers) (Soil Survey Staff, 1993).

Direct Push

Rotary Drill

Hollow stem auger

18-inch (45.72 cm) split barrel with
6-inch (15.24 cm) stainless steel liners

Figure G.3 Collecting soil samples with sampling devices using a drill rig (photos: K. Bessie and N. Harckham).

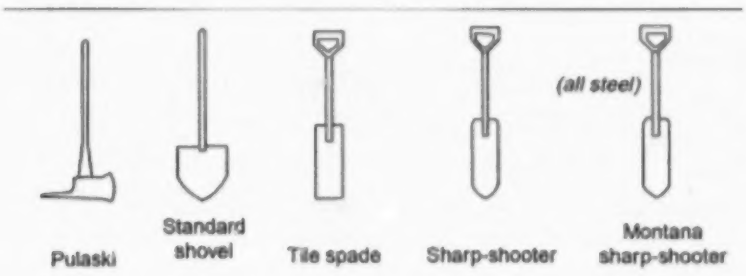


Figure G.4 Select examples of manual field sampling equipment (axes and shovels) (Schoeneberger *et al.*, 2002).

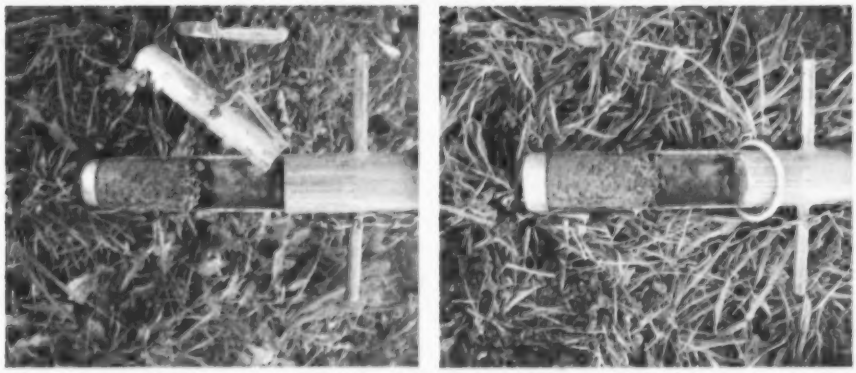


Figure G.5 Split soil corer (photos: J. Römbke).

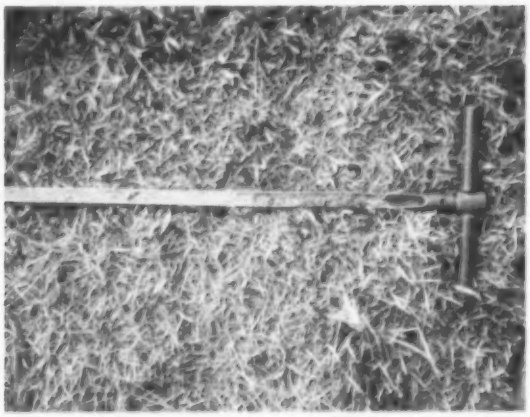


Figure G.6 Pirkhauer corer (photo: J. Römbke).

Appendix H

Soil Sample Containers

A list of the containers recommended for soil samples collected for biological testing is provided in Table H.1.

Table H.1 Containers recommended for soil samples collected for biological testing.

Container	Sample volume (L)	Contamination	Advantages	Disadvantages
High density polyethylene (HDPE) bucket	10 to 20	Inorganics Weathered organics Semi-volatile compounds Reference soils	Widely available Inexpensive Rugged Suitable for long-term storage	Not suitable for volatile organic compounds (VOCs)
Stainless-steel bucket with push-fit lids	5 to 20	Inorganics (except metals) Organics Semi-volatile compounds VOCs	Commercially available Reasonably priced Rugged Suitable for VOCs Suitable for long-term storage	Need specialized equipment to seal buckets
Plastic bag	Up to 60	Inorganics	Usable as a bucket liner for samples contaminated with inorganics (polyethylene bags)	Not rugged
Teflon plastic bag	Up to 60	Inorganics Organics	Chemically inert and solvent resistant to most chemicals. Can be used as a bucket liner or as a sample container by itself	Not rugged
Glass wide-mouthed jars with plastic caps and HDPE lids	0.125 to 2	Inorganics Organics Semi-volatile compounds VOCs	Widely available Inexpensive Suitable for long-term storage	Can only contain very small sample volumes Not rugged
Plastic wide-mouth jars with plastic caps and HDPE lids	0.125 to 4	Inorganics Weathered organics	Widely available Inexpensive Rugged Suitable for long-term storage	Can only contain very small sample volumes Not suitable for non-weathered organics

Appendix H

Soil Sample Containers

A list of the containers recommended for soil samples collected for biological testing is provided in Table H.1.

Table H.1 Containers recommended for soil samples collected for biological testing.

Container	Sample volume (L)	Contamination	Advantages	Disadvantages
High density polyethylene (HDPE) bucket	10 to 20	Inorganics Weathered organics Semi-volatile compounds Reference soils	Widely available Inexpensive Rugged Suitable for long-term storage	Not suitable for volatile organic compounds (VOCs)
Stainless-steel bucket with push-fit lids	5 to 20	Inorganics (except metals) Organics Semi-volatile compounds VOCs	Commercially available Reasonably priced Rugged Suitable for VOCs Suitable for long-term storage	Need specialized equipment to seal buckets
Plastic bag	Up to 60	Inorganics	Usable as a bucket liner for samples contaminated with inorganics (polyethylene bags)	Not rugged
Teflon plastic bag	Up to 60	Inorganics Organics	Chemically inert and solvent resistant to most chemicals. Can be used as a bucket liner or as a sample container by itself	Not rugged
Glass wide-mouthed jars with plastic caps and HDPE lids	0.125 to 2	Inorganics Organics Semi-volatile compounds VOCs	Widely available Inexpensive Suitable for long-term storage	Can only contain very small sample volumes Not rugged
Plastic wide-mouth jars with plastic caps and HDPE lids	0.125 to 4	Inorganics Weathered organics	Widely available Inexpensive Rugged Suitable for long-term storage	Can only contain very small sample volumes Not suitable for non-weathered organics

CHAIN OF CUSTODY RECORD

Shipping Address: LABORATORY NAME
LABORATORY ADDRESS

Telephone: **Fax:**

P.O. Number:
Field Sampler Name (print):
Signature:
Affiliation:
Sample Prep & Storage (before shipping):
Custody Relinquished by:
Date/Time Shipped:

Client:
Phone:
Fax:
Contact:

Sample Identification					Tests Requested			Sample Method and Volume		
Date Collected (yyyy-mm-dd)	Time Collected (e.g. 14:30, 24 hr clock)		Laboratory Sample Number	Temp. on arrival (°C)				Grab Sample	Composite Sample	# of Containers and Volume (eg. 2 x 20L, 3 x 10L)

For Lab Use Only	
Received By:	
Date:	
Time:	
Storage Location:	
Storage Temp (°C):	

Please list any special requests or instructions:

	Name (print)	Signature	Date/ Time
Relinquished by:			
Received by:			
Relinquished by:			
Received by:			
Relinquished by:			
Received by:			

Statistical Tables

Table J.1 The normal distribution[59] (probability Z > value in margin)

Example: Probability Z > 1.96 = 0.025

Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

[59] Generated using function pnorm, R Development Core Team (2008).

Table J.2 Student's *t*-distribution[60] (df = 1- 30).

	Probability of a larger absolute value of *t* (two-sided)								
df	**0.5**	**0.4**	**0.3**	**0.2**	**0.1**	**0.05**	**0.02**	**0.01**	**0.001**
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
	0.25	**0.2**	**0.15**	**0.1**	**0.05**	**0.025**	**0.01**	**0.005**	**5E-04**
	Probability of a larger positive value of *t* (one-sided)								

[60] Generated using function qt, R Development Core Team (2008).

Table J.3 Student's *t*-distribution[61] (df = 31-60).

	Probability of a larger absolute value of *t* (two-sided)								
df	**0.5**	**0.4**	**0.3**	**0.2**	**0.1**	**0.05**	**0.02**	**0.01**	**0.001**
31	0.682	0.853	1.054	1.309	1.696	2.040	2.453	2.744	3.633
32	0.682	0.853	1.054	1.309	1.694	2.037	2.449	2.738	3.622
33	0.682	0.853	1.053	1.308	1.692	2.035	2.445	2.733	3.611
34	0.682	0.852	1.052	1.307	1.691	2.032	2.441	2.728	3.601
35	0.682	0.852	1.052	1.306	1.690	2.030	2.438	2.724	3.591
36	0.681	0.852	1.052	1.306	1.688	2.028	2.434	2.719	3.582
37	0.681	0.851	1.051	1.305	1.687	2.026	2.431	2.715	3.574
38	0.681	0.851	1.051	1.304	1.686	2.024	2.429	2.712	3.566
39	0.681	0.851	1.050	1.304	1.685	2.023	2.426	2.708	3.558
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
41	0.681	0.850	1.050	1.303	1.683	2.020	2.421	2.701	3.544
42	0.680	0.850	1.049	1.302	1.682	2.018	2.418	2.698	3.538
43	0.680	0.850	1.049	1.302	1.681	2.017	2.416	2.695	3.532
44	0.680	0.850	1.049	1.301	1.680	2.015	2.414	2.692	3.526
45	0.680	0.850	1.049	1.301	1.679	2.014	2.412	2.690	3.520
46	0.680	0.850	1.048	1.300	1.679	2.013	2.410	2.687	3.515
47	0.680	0.849	1.048	1.300	1.678	2.012	2.408	2.685	3.510
48	0.680	0.849	1.048	1.299	1.677	2.011	2.407	2.682	3.505
49	0.680	0.849	1.048	1.299	1.677	2.010	2.405	2.680	3.500
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	3.496
51	0.679	0.849	1.047	1.298	1.675	2.008	2.402	2.676	3.492
52	0.679	0.849	1.047	1.298	1.675	2.007	2.400	2.674	3.488
53	0.679	0.848	1.047	1.298	1.674	2.006	2.399	2.672	3.484
54	0.679	0.848	1.046	1.297	1.674	2.005	2.397	2.670	3.480
55	0.679	0.848	1.046	1.297	1.673	2.004	2.396	2.668	3.476
56	0.679	0.848	1.046	1.297	1.673	2.003	2.395	2.667	3.473
57	0.679	0.848	1.046	1.297	1.672	2.002	2.394	2.665	3.470
58	0.679	0.848	1.046	1.296	1.672	2.002	2.392	2.663	3.466
59	0.679	0.848	1.046	1.296	1.671	2.001	2.391	2.662	3.463
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.460
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291
	0.25	**0.2**	**0.15**	**0.1**	**0.05**	**0.025**	**0.01**	**0.005**	**5E-04**
	Probability of a larger positive value of *t* (one-sided)								

[61] Generated using function qt, R Development Core Team (2008).

www.ec.gc.ca

Additional information can be obtained at:

Environment Canada
Inquiry Centre
10 Wellington Street, 23rd Floor
Gatineau QC K1A 0H3
Telephone: 1-800-668-6767 (in Canada only) or 819-997-2800
Fax: 819-994-1412
TTY: 819-994-0736
Email: enviroinfo@ec.gc.ca